

# Explanation of Stagnation at Points that are not Local Optima in Particle Swarm Optimization by Potential Analysis

Alexander Rass      Manuel Schmitt      Rolf Wanka

Department of Computer Science  
University of Erlangen-Nuremberg, Germany  
{alexander.rass, manuel.schmitt, rolf.wanka}@fau.de

May 1, 2015

## Abstract

Particle Swarm Optimization (PSO) is a nature-inspired meta-heuristic for solving continuous optimization problems. In [16, 18], the potential of the particles of a swarm has been used to show that slightly modified PSO guarantees convergence to local optima. Here we show that under specific circumstances the unmodified PSO, even with swarm parameters known (from the literature) to be “good”, almost surely does not yield convergence to a local optimum is provided. This undesirable phenomenon is called stagnation. For this purpose, the particles’ potential in each dimension is analyzed mathematically. Additionally, some reasonable assumptions on the behavior of the particles’ potential are made. Depending on the objective function and, interestingly, the number of particles, the potential in some dimensions may decrease much faster than in other dimensions. Therefore, these dimensions lose relevance, i. e., the contribution of their entries to the decisions about attractor updates becomes insignificant and, with positive probability, they never regain relevance. If Brownian Motion is assumed to be an approximation of the time-dependent drop of potential, practical, i. e., large values for this probability are calculated. Finally, on chosen multidimensional polynomials of degree two, experiments are provided showing that the required circumstances occur quite frequently. Furthermore, experiments are provided showing that even when the very simple sphere function is processed the described stagnation phenomenon occurs. Consequently, unmodified PSO does not converge to any local optimum of the chosen functions for tested parameter settings.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Definitions</b>	<b>5</b>
<b>3</b>	<b>Stagnation Analysis of PSO</b>	<b>8</b>
3.1	Theoretical Tools . . . . .	8
3.2	Potential and Stagnation Phases . . . . .	12
3.3	Unlimited Stagnation Phases . . . . .	14
3.4	Convergence to Non-Optimal Points . . . . .	20
<b>4</b>	<b>Experimental Results</b>	<b>27</b>
4.1	Used Software . . . . .	27
4.2	Expectation of increments for various scenarios . . . . .	27
4.3	Verification of the Assumption of Separation . . . . .	39
4.4	Experimental Verification of the Assumption of Independence . . . . .	43
4.5	Brownian Motion as Approximation for Accumulated Sums . . . . .	47
4.6	Duration of Phases of Type $PH_X$ . . . . .	49
4.7	Swarm Converges . . . . .	51
4.8	Empirical Distributions of Stopping Times $\alpha_i$ and $\beta_i$ and of Logarithmic Potential . . . . .	53
4.9	Long Time Experiments . . . . .	57
<b>5</b>	<b>Conclusion</b>	<b>58</b>
<b>A</b>	<b>Data of Experiment 1</b>	<b>62</b>
<b>B</b>	<b>Data of Experiment 3</b>	<b>64</b>

# 1 Introduction

Particle swarm optimization (PSO), introduced by Kennedy and Eberhart [9, 6], is a very popular nature-inspired meta-heuristic for solving continuous optimization problems. Fields of very successful application are, among many others, Biomedical Image Processing [21], Geosciences [11], and Materials Science [13], where the continuous objective function on a multi-dimensional domain is not given in a closed form, but by a “black box”. The popularity of the PSO framework is due to the fact that on the one hand it can be realized and, if necessary, adapted to further needs easily, but on the other hand shows in experiments good performance results with respect to the quality of the obtained solution and the speed needed to obtain it. A thorough discussion of PSO can be found in [12].

To be precise, let an objective function  $f : \mathbb{R}^D \rightarrow \mathbb{R}$  on a  $D$ -dimensional domain be given that (w.l.o.g.) has to be minimized. A population of *particles*, each consisting of a position (the candidate for a solution), a velocity and a local attractor, moves through the search space  $\mathbb{R}^D$ . The local attractor of a particle is the best position with respect to  $f$  this particle has encountered so far. The best of all local attractors is the global attractor. The movement of a particle is governed by so-called movement equations that depend on both the particle’s velocity and its two attractors and on some additional fixed algorithm parameters. The pseudo code of the PSO approach is visualized in Algorithm 1. Additionally, Definition 1 captures the PSO behavior mathematically as a *stochastic process*. The population in motion is called the *swarm*.

There are guidelines known for the “good” choice of the fixed parameters that control the impact of the current velocity and the attractors on the updated velocity of a particle ([20, 7]) such that the swarm provably converges to a particular point in the search space (under some reasonable assumptions). However, the point of convergence is not necessarily a global optimum. Local optima might also be considered acceptable, but unfortunately it is possible that the point of convergence is not even a local optimum. In the latter case, one says that the swarm *stagnates*. Examples are presented in Sec. 4. E. g., with established *good* parameter settings, stagnation can be observed if 3 particles work on the 10-dimensional, sphere function. In [10], Lehre and Witt show that there are non-trivial bad parameter settings and initial configurations such that PSO possibly converges at arbitrary points when processing the one-dimensional sphere function. However, their result is presented for populations of 1 and 2 particles only. Additionally, their parameter settings considerably deviate from those generally considered as good ([3, 4, 20]).

In [16, 18], the notion of the potential of the swarm has been introduced. This potential has been used to prove that in the one-dimensional search space the swarm almost surely (in the mathematical sense) finds a local optimum. If  $D \geq 2$ , the movement equation has been adapted slightly to avoid that the swarm’s potential drops too close to 0, and hence avoiding stagnation at a non-optimal position. Consequently, this version of PSO almost surely finds local optima.

A comprehensive overview on theoretical results concerning PSO can be found in [15].

The phenomenon of convergence to a point that is not a local optimum, has to the best of our knowledge not yet been formally investigated in a setting which is not generally restricted to a number of particles or specific parameters for the PSO. In this paper, reasons for the phenomenon of stagnation are given. For it, the notion of the potential is modified such that it is now aware of how the particles experience the objective function  $f$ . A theoretical model is provided which on the one hand uses reasonable assumptions and on the other hand provides a basis to mathematically prove that the swarm almost surely stagnates. The model captures the observation that during the execution of PSO on objective functions  $f$  the contribution of some dimensions to the potential

decreases exponentially faster than the contribution of other dimensions. This does not generally imply that PSO faces the problem of stagnation, but if the model is applicable, which is often the case when few particles are used, then this model supplies indications whether in a specific setting stagnation is present or not. The assumptions which need to be applied for the model will be justified in the experimental Section 4.

The first model is used to prove the statement, that finally the swarm stagnates indefinitely almost surely. In that model we can prove that the conditional probability that stagnation remains indefinitely if stagnation emerges at some time  $T$  is positive. If the time-dependent drop of potential is additionally approximated by a Brownian Motion with drift, then the described probability can be calculated by an explicit formula. The values obtained from that formula coincide very well with empirically measured probabilities in experiments. If stagnation remains indefinitely mainly poor solutions will be returned by PSO.

Experimentally it is shown that the described separation of potential indeed occurs when PSO is run with certain numbers of particles on some popular functions from the CEC benchmark set [19] and an additional function.

This paper is organized as follows: After the formal definition of the PSO process in Section 2, in Section 3 the new notion of potential, results of experiments and the proof that under realistic assumptions PSO almost surely stagnates, are presented. In Section 4, experimental analysis of three benchmark functions from the CEC benchmark set [19] and an additional function is presented.

## 2 Definitions

First the model which is going to be used for our analysis of the PSO algorithm is presented. Algorithm 1 represents the pseudo code of the classical PSO algorithm. No bound handling strategies are investigated, because they have almost no influence on the convergence if the swarm is converging to a point not on the boundaries. Similar to [16, 18], the model describes the positions of the particles, the velocities and the global and local attractors also as real-valued stochastic processes. Basic mathematical tools from probability theory, which are needed for this analysis can be found in, e. g., [5].

**Definition 1** (Classical PSO process). *A swarm of  $N$  particles moves through the  $D$ -dimensional search space  $\mathbb{R}^D$ . Let  $f : \mathbb{R}^D \rightarrow \mathbb{R}$  be the objective function. At each time  $t \in \mathbb{N}$ , each particle  $n$  has a position  $X_t^n \in \mathbb{R}^D$ , a velocity  $V_t^n \in \mathbb{R}^D$  and a local attractor  $L_t^n \in \mathbb{R}^D$ , the best position particle  $n$  has visited until time  $t$ . Additionally, the swarm shares its global attractor  $G_t^n \in \mathbb{R}^D$ , storing the best position any particle has visited until the  $t$ 'th step of particle  $n$ , i. e., since each particle's update of the global attractor is immediately visible for the next particle, it can have several values in the same iteration. When necessary,  $X_t^{n,d}$  is written for the  $d$ 'th component of  $X_t^n$  ( $V_t^{n,d}$  etc. analogously). With a given distribution for  $(X_0, V_0)$ ,  $(X_{t+1}, V_{t+1}, L_{t+1}, G_{t+1})$  is determined by the following recursive equations that are called the movement equations:*

$$\begin{aligned} L_0^n &:= X_0^n \text{ for } 1 \leq n \leq N, \\ G_t^n &:= \operatorname{argmin}_{x \in \bigcup_{i=1}^{n-1} \{L_{t+1}^i\} \cup \bigcup_{i=n}^N \{L_t^i\}} f(x) \text{ for } t \geq 0, 1 \leq n \leq N, \\ V_{t+1}^n &:= \chi \cdot V_t^n + c_1 \cdot r_t^n \odot (L_t^n - X_t^n) + c_2 \cdot s_t^n \odot (G_t^n - X_t^n) \text{ for } t \geq 0, 1 \leq n \leq N, \\ X_{t+1}^n &:= X_t^n + V_{t+1}^n \text{ for } t \geq 0, 1 \leq n \leq N, \\ L_{t+1}^n &:= \operatorname{argmin}_{x \in \{X_{t+1}^n, L_t^n\}} f(x) \text{ for } t \geq 0, 1 \leq n \leq N. \end{aligned}$$

Here,  $\chi$ ,  $c_1$  and  $c_2$  are some positive constants called the fixed parameters of the swarm, and  $r_t^n$ ,  $s_t^n$  are uniformly distributed over  $[0, 1]^D$  and all independent.  $\odot$  is meant as the item-wise product of two vectors. The result is then a vector as well.

The underlying probability space is called  $(\Omega, \mathcal{A}, P)$ . The  $\sigma$ -Algebra  $\mathcal{A}$  is equal to the generated  $\sigma$ -Algebra of the family of  $\sigma$ -Algebras  $\{\mathcal{A}_t\}_{t \in \mathbb{N}}$ , to which  $(X_t, V_t, L_t)_{t \in \mathbb{N}}$  is adapted.

One can think of  $\mathcal{A}_t$  as the mathematical object carrying the information known at time  $t$ . Mainly  $\mathcal{A}_t$  refers to the product  $\sigma$ -algebra for the random variables  $r_t^n$  and  $s_t^n$ , but we will introduce further random variables, which will be measurable by some  $\mathcal{A}_t$  or  $\mathcal{A}$  as well, to expand our model. If after the  $t$ 'th step the process is stopped, the *solution* is  $G_t^1$ .  $G_t^1$  is  $\mathcal{A}_t$ -measurable because it is the argmin of the local attractors  $L_t^1, \dots, L_t^N$ , whereas  $G_t$  is not  $\mathcal{A}_t$ -measurable as it partially depends on  $L_{t+1}$ . In [1], one can find a comprehensive introduction into  $\sigma$ -algebras and filtrations. Informally,  $\mathcal{A}_t$  captures the information of the underlying stochastic process, that is available at time  $t$ , i. e., a random variable is  $\mathcal{A}_t$ -measurable, iff its value is determined at time  $t$  and for a random variable  $V$ ,  $E[V \mid \mathcal{A}_t]$  is the expectation of  $V$  under taking the information of every time  $t' \leq t$  into account.  $E[V \mid \mathcal{A}_t]$  is an  $\mathcal{A}_t$ -measurable random variable, e. g., if  $V$  is  $\mathcal{A}_t$ -measurable then  $E[V \mid \mathcal{A}_t] = V$ .

The update process of a particle is visualized in Figure 1. As specified in the movement equations of Definition 1, the new velocity  $V_{t+1}^n$  of the particle  $n$  consists of a fraction of the old particle

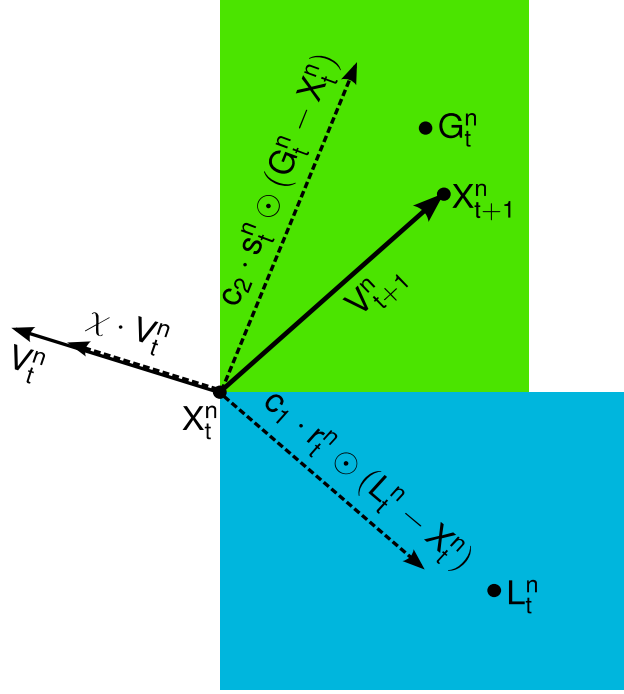


Figure 1: Update process of a particle with  $\chi = 0.72984$  and  $c_1 = c_2 = 1.496172$ : Position  $X$ , velocity  $V$ , local attractor  $L$  and global attractor  $G$

without any randomness  $\chi \cdot V_t^n$ , a randomized part of the vector from the current position to the local attractor  $c_1 \cdot r_t^n \odot (L_t^n - X_t^n)$  and a randomized part of the vector from the current position to the global attractor  $c_2 \cdot s_t^n \odot (G_t^n - X_t^n)$ . With PSO parameters  $\chi = 0.72984$  and  $c_1 = c_2 = 1.496172$ , as proposed in [4], the terminal point of the randomized vector from the current position to the local attractor is sampled uniformly in the blue area of Figure 1 and the terminal point of the randomized vector from the current position to the global attractor is sampled uniformly in the green area of Figure 1. A possible assignment for the three vectors is visualized by the dashed vectors. The accumulation of the three parts yields the new velocity  $V_{t+1}^n$  and the new position  $X_{t+1}^n$  can be reached by moving  $X_t^n$  by  $V_{t+1}^n$ . After the update of the velocity and the position, the local and global attractor could be updated. If also this step is executed, then it can be proceeded with the next particle. Algorithm 1 determines the respective approach with pseudo code.

---

**Algorithm 1:** Classical PSO

---

```
input : objective function  $f : \mathbb{R}^D \rightarrow \mathbb{R}$ 
output: an optimized position  $G \in \mathbb{R}^D$  for the objective function  $f$ 

/* initialize all  $N$  particles: */
/* the positions  $X \in (\mathbb{R}^D)^N$  and the velocities  $V \in (\mathbb{R}^D)^N$  */
1  $(X, V) := \text{getInitialPositionsAndVelocities}();$ 
/* the local attractors  $L \in (\mathbb{R}^D)^N$  */
2  $L := X;$ 
/* initialize the global attractor  $G \in \mathbb{R}^D$  */
3  $G := \text{argmin}_{x \in \{L[1], \dots, L[N]\}} f(x);$ 
4 while termination criterion not fulfilled do
5   for  $n := 1$  to  $N$  do
6     for  $d := 1$  to  $D$  do
7       /* update  $d$ 'th velocity coordinate of  $n$ 'th particle */
8       /*  $\text{rand}(a, b)$  supplies a uniform random value in  $[a, b]$  */
9        $V[n][d] := \chi \cdot V[n][d] +$ 
10       $+c_1 \cdot \text{rand}(0.0, 1.0) \cdot (L[n][d] - X[n][d])$ 
11       $+c_2 \cdot \text{rand}(0.0, 1.0) \cdot (G[d] - X[n][d])$ 
12     end
13     /* update position of  $n$ 'th particle */
14      $X[n] := X[n] + V[n];$ 
15     /* update local attractor of  $n$ 'th particle */
16     if  $f(X[n]) < f(L[n])$  then
17        $L[n] := X[n];$ 
18     end
19     /* update global attractor */
20     if  $f(X[n]) < f(G)$  then
21        $G := X[n];$ 
22     end
23   end
24 end
25 return  $G$  ;
```

---

### 3 Stagnation Analysis of PSO

In this chapter some assumptions are made, such that theoretical analysis can be done. With these assumptions theoretical proofs are provided, which state that PSO does not reach local optima almost surely if the number of particles is too small.

#### 3.1 Theoretical Tools

First of all some theoretical tools are introduced to make later proofs more manageable. The following technical, well known lemma, which will be used in the proof of Lemma 2, gives sufficient conditions for an infinite product of probabilities to have a positive value.

**Lemma 1.** *If  $0 \leq a_i < 1$  for every  $i \in \mathbb{N}$  and  $\sum_{i=1}^{\infty} a_i < \infty$ , then*

$$\prod_{i=1}^{\infty} (1 - a_i) > 0 .$$

*Proof.*  $\prod_{i=1}^{\infty} (1 - a_i) > 0$  iff  $\exists n_0 \in \mathbb{N} : \prod_{i=n_0}^{\infty} (1 - a_i) > 0$ . Choose  $n_0$  sufficiently large such that  $\sum_{i=n_0}^{\infty} a_i < 1$ . It follows that  $\prod_{i=n_0}^{\infty} (1 - a_i) \geq 1 - \sum_{i=n_0}^{\infty} a_i > 0$ .  $\square$

In the following, the Markov's inequality is applied to prove that any finite sum of independent, identically distributed random variables, with negative expectation and finite first six moments each, stays below zero with probability  $> 0$ .

**Lemma 2.** *Let  $(I_t)_{t \in \mathbb{N}}$  be independent identically distributed (i.i.d.) random variables with expected value  $\mu^*$ , let  $\mu < 0$ ,  $M > 0$  and  $p_0 > 0$  be constants. If  $\mu^* \leq \mu$ ,  $E[(I_t - \mu^*)^i] \leq M$  for all  $i \in \{1, \dots, 6\}$  and  $P(I_t \leq 0) \geq p_0$ , then:*

- $P(\sum_{i=0}^{t-1} I_i > 0) \leq \frac{C}{t^3}$  for a constant  $C > 0$ , such that  $C := C(\mu, M)$ , and for all  $t > 0$ ,
- $P(\forall t \in \mathbb{N} : \sum_{i=0}^{t-1} I_i \leq 0) \geq p$  for a constant  $p > 0$ , such that  $p := p(\mu, M, p_0)$ .

*Proof.* As  $E[I_0 - \mu^*] = 0$  and  $I_t$  are i.i.d., we get

$$\begin{aligned} E\left[\left(\sum_{i=0}^{t-1} (I_i - \mu^*)\right)^6\right] &= t \cdot E[(I_0 - \mu^*)^6] + 10t(t-1) \cdot E[(I_0 - \mu^*)^3]^2 \\ &\quad + 15t(t-1)E[(I_0 - \mu^*)^4]E[(I_0 - \mu^*)^2] + 15t(t-1)(t-2)E[(I_0 - \mu^*)^2]^3 \\ &\leq \tilde{C} \cdot t^3, \end{aligned}$$

where

$$\begin{aligned} \tilde{C} &:= \tilde{C}(M) := M + 10 \cdot M^2 + 15 \cdot M^2 + 15 \cdot M^3 \\ &\geq E[(I_0 - \mu^*)^6] + 10E[(I_0 - \mu^*)^3]^2 + 15E[(I_0 - \mu^*)^4]E[(I_0 - \mu^*)^2] + 15E[(I_0 - \mu^*)^2]^3, \end{aligned}$$

because every product which contains  $E[(I_t - \mu^*)^1]$  as a factor is zero and therefore only the exponent six, the pair of exponents two and four, the pair of exponents three and three and the



tuple of exponents two, two and two remain. With the Markov's inequality on the sixth power of the centralized sum, which is a positive random variable, we get

$$\begin{aligned} \mathbb{P}\left(\sum_{\tilde{t}=0}^{t-1} I_{\tilde{t}} > 0\right) &\leq \mathbb{P}\left(\sum_{\tilde{t}=0}^{t-1} (I_{\tilde{t}} - \mu^*) \geq -\mu^* t\right) \\ &\stackrel{\text{Markov}}{\leq} \frac{\mathbb{E}[(\sum_{\tilde{t}=0}^{t-1} (I_{\tilde{t}} - \mu^*))^6]}{(\mu^* t)^6} \leq \frac{\tilde{C} \cdot t^3}{(\mu^* t)^6} \stackrel{|\mu^*| \geq |\mu|}{\leq} \frac{\tilde{C}}{\mu^6 t^3} \leq \frac{C}{t^3} \end{aligned}$$

where  $C := C(\mu, M) := \tilde{C}(M)/\mu^6$ . Those inequalities prove the first part of Lemma 2. To complete the proof further inequalities are needed, which will be received with conditional probabilities. Let  $\tilde{t} < \tilde{T}$  be a non negative integer.

$$\begin{aligned} &\mathbb{P}\left(\sum_{t=0}^{\tilde{T}} I_t \leq 0 \mid \sum_{t=0}^{\tilde{t}} I_t \leq 0 \wedge \forall t' < \tilde{t} : \sum_{t=0}^{t'} I_t \leq 0\right) \\ &\geq \mathbb{P}\left(\sum_{t=\tilde{t}+1}^{\tilde{T}} I_t \leq 0 \mid \sum_{t=0}^{\tilde{t}} I_t \leq 0 \wedge \forall t' < \tilde{t} : \sum_{t=0}^{t'} I_t \leq 0\right) \\ &= \mathbb{P}\left(\sum_{t=\tilde{t}+1}^{\tilde{T}} I_t \leq 0 \mid \sum_{t=0}^{\tilde{t}} I_t > 0 \wedge \forall t' < \tilde{t} : \sum_{t=0}^{t'} I_t \leq 0\right) \\ &\geq \mathbb{P}\left(\sum_{t=0}^{\tilde{T}} I_t \leq 0 \mid \sum_{t=0}^{\tilde{t}} I_t > 0 \wedge \forall t' < \tilde{t} : \sum_{t=0}^{t'} I_t \leq 0\right). \end{aligned}$$

As

$$\mathbb{P}\left(\sum_{t=0}^{\tilde{T}} I_t \leq 0 \mid \forall t' < \tilde{t} : \sum_{t=0}^{t'} I_t \leq 0\right)$$

is a convex combination of the first and the last term of the recent series of inequalities, it is less or equal to the greater value

$$\mathbb{P}\left(\sum_{t=0}^{\tilde{T}} I_t \leq 0 \mid \sum_{t=0}^{\tilde{t}} I_t \leq 0 \wedge \forall t' < \tilde{t} : \sum_{t=0}^{t'} I_t \leq 0\right)$$

and by induction

$$\mathbb{P}\left(\sum_{t=0}^{\tilde{T}} I_t \leq 0 \mid \forall t' < \tilde{T} : \sum_{t=0}^{t'} I_t \leq 0\right) \geq \mathbb{P}\left(\sum_{t=0}^{\tilde{T}} I_t \leq 0\right).$$

Summing up all these probabilities leads to a sum, which is bounded:

$$\sum_{t=1}^{\infty} \mathbb{P}\left(\sum_{\tilde{t}=0}^{t-1} I_{\tilde{t}} > 0\right) \leq \sum_{t=1}^{\infty} \frac{C}{t^3} < \infty \quad (1)$$

Since  $\mathbb{P}(I_t \leq 0) \geq p_0 > 0$ , it follows that

$$\mathbb{P}\left(\forall t \leq T : \sum_{\tilde{t}=0}^{t-1} I_{\tilde{t}} \leq 0\right) \geq \mathbb{P}(\forall t \leq T : I_t \leq 0) \geq p_0^T > 0.$$

Therefore the probability to reach no positive value is positive for fixed finite times. As mentioned in equation 1 the infinite sum with the probabilities, which specify the probability to remain non-positive at a specific time, is bounded and therefore Lemma 1 is applicable. Altogether we have

$$\begin{aligned}
\mathbb{P}\left(\forall t \in \mathbb{N} : \sum_{\tilde{t}=0}^{t-1} I_{\tilde{t}} \leq 0\right) &= \lim_{t \rightarrow \infty} \mathbb{P}\left(\forall t' \leq t : \sum_{\tilde{t}=0}^{t'-1} I_{\tilde{t}} \leq 0\right) \\
&= \mathbb{P}(I_0 \leq 0) \prod_{\tilde{t}=2}^{\infty} \mathbb{P}\left(\sum_{t=0}^{\tilde{t}-1} I_t \leq 0 \mid \forall t' < \tilde{t} : \sum_{t=0}^{t'-1} I_t \leq 0\right) \\
&\geq \prod_{\tilde{t}=1}^{\infty} \mathbb{P}\left(\sum_{t=0}^{\tilde{t}-1} I_t \leq 0\right) \\
&\geq \underbrace{\prod_{\tilde{t}=1}^{\infty} \left(1 - \min\left(1 - p_0, \frac{C}{\tilde{t}^3}\right)\right)}_{>0 \text{ with Lemma 1}} =: p > 0.
\end{aligned}$$

Obviously  $p$  depends only on  $\mu$ ,  $M$  and  $p_0$  because in the last expression only  $C := C(\mu, M)$  and  $p_0$  appear.  $\square$

The next lemma provides some estimation for the moments of two random variables if a bound for the sixth moment of their sum is provided.

**Lemma 3.** *Let  $A$  and  $B$  be independent random variables with finite first six moments and expectation zero. There exists a fixed function  $h : \mathbb{R} \rightarrow \mathbb{R}$  such that for each  $i \in \{1, \dots, 6\}$*

- $|\mathbb{E}[(A + B)^i]| \leq h(M)$
- $|\mathbb{E}[A^i]| \leq h(M)$
- $|\mathbb{E}[B^i]| \leq h(M)$

if  $\mathbb{E}[(A + B)^6] \leq M$ .

*Proof.* For each  $i \in \{1, \dots, 6\}$  we define fixed functions  $h_{1,i} : \mathbb{R} \rightarrow \mathbb{R}$ , and  $h_{2,i} : \mathbb{R} \rightarrow \mathbb{R}$  such that

- $|\mathbb{E}[(A + B)^i]| \leq h_{1,i}(M)$ ,
- $|\mathbb{E}[A^i]| \leq h_{2,i}(M)$  and
- $|\mathbb{E}[B^i]| \leq h_{2,i}(M)$ ,

if  $\mathbb{E}[(A + B)^6] \leq M$ .

All expected values exist as specified. Only the inequalities without absolute value need to be shown because all evaluations can also be made with  $-A$  instead of  $A$  and  $-B$  instead of  $B$ , which guarantees the inequality for odd  $i$ . For even  $i$  the absolute value has no effect on the equation because the value inside the absolute value is already non-negative. As random variables  $A$  and

$B$  are independent, the expectation of a product can easily be separated. For  $i = 1$  all expected values are zero and therefore the following functions are possible:

$$h_{1,1}(M) := h_{2,1}(M) := 0.$$

$h_{1,i}(M) := M + 1$  for all  $i \in \{2, \dots, 6\}$ , because

$$\begin{aligned} \mathbb{E}[(A+B)^i] &\leq \mathbb{E}[\max(1, (A+B))^i] \\ &\stackrel{i \leq 6}{\leq} \mathbb{E}[\max(1, (A+B)^6)] \\ &\leq \mathbb{E}[1 + (A+B)^6] \\ &\leq 1 + M. \end{aligned} \tag{2}$$

$$\mathbb{E}[(A+B)^4] = \mathbb{E}[A^4] + 4\mathbb{E}[A^3] \underbrace{\mathbb{E}[B^1]}_{=0} + \underbrace{6\mathbb{E}[A^2]\mathbb{E}[B^2]}_{\geq 0} + 4\mathbb{E}[A^1] \underbrace{\mathbb{E}[B^3]}_{=0} + \mathbb{E}[B^4],$$

which leads to

$$\underbrace{\mathbb{E}[A^4]}_{\geq 0} + \underbrace{\mathbb{E}[B^4]}_{\geq 0} \leq \mathbb{E}[(A+B)^4] \leq h_{1,4}(M)$$

and therefore

$$h_{2,4}(M) := h_{1,4}(M) = M + 1$$

is possible. Similar to inequalities 2

$$h_{2,i}(M) := h_{2,4}(M) + 1 = M + 2$$

is received for all  $i \in \{1, 2, 3\}$ . Similar to the case  $i = 4$  we receive

$$\begin{aligned} \underbrace{\mathbb{E}[(A+B)^6]}_{\leq h_{1,6}(M)} &= \mathbb{E}[A^6] + \binom{6}{2} \underbrace{\mathbb{E}[A^4]\mathbb{E}[B^2]}_{\geq 0} + \binom{6}{3} \underbrace{\mathbb{E}[A^3]\mathbb{E}[B^3]}_{\geq -h_{2,3}(M)^2} + \binom{6}{4} \underbrace{\mathbb{E}[A^2]\mathbb{E}[B^4]}_{\geq 0} + \mathbb{E}[B^6] \\ &\Rightarrow \underbrace{\mathbb{E}[A^6]}_{\geq 0} + \underbrace{\mathbb{E}[B^6]}_{\geq 0} \leq h_{1,6}(M) + \binom{6}{3} h_{2,3}(M)^2 =: h_{2,6}(M) \end{aligned}$$

and finally  $h_{2,5}(M) := 1 + h_{2,6}(M)$ .

For each  $M$  the maximal function value of all functions  $h_{i,j}$  is a suitable choice for the function value of  $h$ .

$$h(M) := \max_{i \in \{1,2\}, j \in \{1, \dots, 6\}} h_{i,j}(M)$$

□

### 3.2 Potential and Stagnation Phases

Now the idea of a stagnation measure is introduced, which is a multidimensional extension to the potential used in [16, 17, 18]. For every step, a  $D$ -dimensional vector of potentials is evaluated – one potential value for each dimension. It is intended that the greater the value of such a potential for a single dimension is, the greater is the impact of this dimension on the behavior of the swarm, i.e., the greater is the portion of the change in the function value, which is due to the movement in that dimension. This property is not declared in the definition because it cannot be quantified in a strict way. It will be specified in detail in the Assumptions 2 and 3. Furthermore, the logarithmic potential is defined, which compares the impact of a specific dimension with the maximal impact along all dimensions. The dimension which has currently the highest impact on the swarm has a logarithmic potential of zero and all other dimensions will have no larger logarithmic potential. Since the convergence analysis in [8] implies that the general movement of a converging particle swarm drops exponentially, a logarithmic scale is used and linear decrease is expected.

**Definition 2** (Potential). *Let  $\tilde{\Phi} : \mathbb{R}^{3 \cdot N \cdot D} \rightarrow \mathbb{R}^D$  be a measurable function. Let  $\Delta t$  be a positive integer constant, which will be called the step width.  $\Phi(t, d)$  is defined as*

$$\Phi(t, d) := (\tilde{\Phi}(X_{t \cdot \Delta t}, V_{t \cdot \Delta t}, L_{t \cdot \Delta t}))_d.$$

$\Phi : \mathbb{N}_0 \times \{1, \dots, D\} \rightarrow (\Omega \rightarrow \mathbb{R})$  is a function which evaluates to a random variable for each pair in  $\mathbb{N}_0 \times \{1, \dots, D\}$ .  $\Phi$  is called a potential if  $\Phi(t, d)$  is positive almost surely for all  $t$  and  $d$ . Additionally

$$\Psi(t, d) := \log \left( \Phi(t, d) / \max_{\tilde{d} \in \{1, \dots, D\}} \Phi(t, \tilde{d}) \right) = \log \left( \Phi(t, d) \right) - \log \left( \max_{\tilde{d} \in \{1, \dots, D\}} \Phi(t, \tilde{d}) \right)$$

is called a logarithmic potential and  $(I_{t,d})_{t \in \mathbb{N}}$ , with  $I_{t,d} := \Psi(t+1, d) - \Psi(t, d)$  are called increments of dimension  $d$ .  $\Phi(t, d)$ ,  $\Psi(t, d)$  and  $I_{t-1,d}$  are  $\mathcal{A}_{t \cdot \Delta t}$ -measurable random variables, where  $\mathcal{A}_t$  is the  $\sigma$ -algebra, which is specified in Definition 1.

In this paper the expression  $\log$  is associated with the logarithm of base 2. The step width  $\Delta t$  specifies how the time is scaled in respect to the potential, which means that the potential is only evaluated for PSO configurations  $(X_t, V_t, L_t)$  if  $t$  is a multiple of  $\Delta t$ . The effect of  $\Delta t$  will be explained later. In Section 4 and in all figures a potential is used which will be similar to the item-wise product of the current velocity and the gradient of the function at the current position. More precisely, the following potential will be used:

**Definition 3** (Experimental potential). *The potential is defined as*

$$\Phi(t, d) := \max_{n \in \{1, \dots, N\}} |f(X_{\Delta t \cdot t}^n) - f(\tilde{X}_{\Delta t \cdot t}^{n,d})|$$

$$\text{where } (\tilde{X}_t^{n,d})_{\tilde{d}} := \begin{cases} X_t^{n,\tilde{d}} + V_t^{n,\tilde{d}}, & \text{if } \tilde{d} = d, \\ X_t^{n,\tilde{d}}, & \text{otherwise} \end{cases}$$

and  $f$  is the objective function which should be optimized.

$\tilde{X}_t^{n,d}$  represents the position of a particle if only a step in dimension  $d$  is done. The deviation in one dimension is defined as a limit of a difference quotient. We get this difference quotient by

dividing the potential by the velocity. If the absolute value of the velocity becomes very small we get something similar to the item-wise product of the current velocity and the gradient as already mentioned. If the absolute value of the gradient in some dimension tends to zero then higher deviations influence the potential. It might be possible to use other potentials but for the objective functions, which are considered, this potential is sufficient. Mainly there are two prerequisites which need to be fulfilled such that this potential is suitable. Firstly, the velocities need to be nonzero almost surely. If not all particles are initialized at the same point, then the velocities will not encounter a value of zero almost surely. Even for zero velocity initialization this is almost surely true, if the evaluation of the potential starts at the second step and standard position initialization is used. Secondly, the objective function needs to fulfill the property that there exists no set of positive measure such that each point evaluates to the same function value. For most continuous functions which have no plateaus this prerequisite is achieved. Therefore this potential is positive for such functions almost surely if a standard PSO initialization is used.

By means of the logarithmic potential, a stagnation phase can be defined. If the logarithmic potential of a dimension becomes very low, then there are other dimensions which have currently much more impact on the swarm behavior. Finally, a dimension which has very low logarithmic potential for a long period of time will be far away from an optimized position in this dimension. To specify this activity more in detail, the following definition is stated:

**Definition 4** ( $(\Delta t, \Phi, N_0, c_0, c_s)$ -Stagnation phase). *Let  $\Delta t$  be a constant step width,  $\Phi$  a potential with the associated logarithmic potentials  $\Psi$  and their increments  $I_{t,d}$ ,  $N_0$  a positive constant integer less than  $D$  and  $c_0 \leq c_s < 0$  negative constants. The following stopping times are defined:*

$$\beta_{-1} := 0$$

and inductively for all  $i \geq 0$ :

$$\alpha_i := \Delta t \cdot \inf \left\{ t \geq \frac{\beta_{i-1}}{\Delta t} : |\{d \in \{1, \dots, D\} : \Psi(t, d) \leq c_0\}| \geq N_0 \right\}$$

$$\beta_i := \Delta t \cdot \inf \left\{ t \geq \frac{\alpha_i}{\Delta t} : |\{d \in \{1, \dots, D\} : \Psi(\frac{\alpha_i}{\Delta t}, d) \leq c_0 \wedge \max_{\frac{\alpha_i}{\Delta t} \leq t' \leq t} \Psi(t', d) \leq c_s\}| < N_0 \right\}$$

If  $\alpha_i$  is finite, then it is said that the  $i$ 'th  $(\Delta t, \Phi, N_0, c_0, c_s)$ -stagnation phase starts at time  $\alpha_i$ . If  $\beta_i$  is finite too, then it is said that the  $i$ 'th  $(\Delta t, \Phi, N_0, c_0, c_s)$ -stagnation phase ends at time  $\beta_i$ . If  $\alpha_i$  is finite but  $\beta_i$  is not finite then the  $i$ 'th  $(\Delta t, \Phi, N_0, c_0, c_s)$ -stagnation phase does not end. The event that a  $(\Delta t, \Phi, N_0, c_0, c_s)$ -stagnation phase starts at time  $T_0 \Delta t$  is defined as  $\{\exists i \in \mathbb{N}_0 : \alpha_i = T_0 \Delta t\}$ .

$N_0$  defines some minimal number of dimensions, which stagnate during the complete phase, i. e., the logarithmic potential in these dimensions is at most  $c_s$  during the stagnation phase.  $c_0$  defines a starting safety distance to the highest potential and  $c_s$  defines a permanent safety distance to the highest potential such that dimensions which stay below that bound remain insignificant. It is intended that it should not happen that stagnating dimensions, i. e., dimensions with low logarithmic potential, influence the attractors, because the swarm behaves differently if additional dimensions influence the attractors. To ensure that stagnating dimensions have almost no influence on the swarm, the logarithmic potential of stagnating dimensions need to be the safety distance apart of the logarithmic potential of dimension with most influence. The bigger this safety distance is the more reasonable is the assumption of actual insignificance of stagnating dimensions. It

might happen that  $\alpha_i = \beta_{i-1}$ . For example the set  $\{1, \dots, N_0\}$  can be the set of dimensions, which have a logarithmic potential of at most  $c_0$  at the start of a stagnation phase. During the stagnation phase the logarithmic potential of dimension  $(N_0 + 1)$  can become less than  $c_0$ . If the logarithmic potential of dimensions  $2, \dots, N_0$  stay below  $c_0$  and the logarithmic potential of dimension 1 increases to a value larger than  $c_s$ , then this stagnation phase ends and the next stagnation phase immediately starts because all dimensions in  $\{2, \dots, N_0 + 1\}$  have a logarithmic potential of less than  $c_0$ . Therefore it can happen that a stagnation phase starts immediately after another stagnation phase has ended, i. e.,  $\alpha_i = \beta_{i-1}$ . Furthermore, stagnation phases will not end at the same time as they start, i. e.,  $\alpha_i > \beta_i$ , because  $c_0 \leq c_s$  and therefore the size of the set of dimensions with low logarithmic potential is at least  $N_0$  at the beginning.

Dimensions with low logarithmic potential have low influence on the change of the value of the objective function during one step compared to the other dimensions. This is obvious in respect to the experimental potential, which is introduced in Definition 3, because the  $d$ 'th value represents the change of the function value if we do a step in dimension  $d$  and leave the other dimensions as they are. If the logarithmic potential of a dimension  $d$  is low for some period of time, this dimension is called stagnating dimension. Stagnating dimensions are called stagnating, because their small impact on the change in the value of the objective function results in a small impact on the decision whether a new position is a local attractor or a global attractor. Therefore the decision whether a new position is an attractor does hardly depend on the movement in the stagnating dimensions.

### 3.3 Unlimited Stagnation Phases

**Definition 5.** Let  $\Theta$  be a subset of probability distributions on  $\mathbb{R}$ . Let  $\Delta t$  be the constant step width. Let  $(B_{t,\Gamma_B})_{t \in \mathbb{N}, \Gamma_B \in \Theta}$  and  $(J_{t,d,\Gamma_J})_{t \in \mathbb{N}, d \in \{1, \dots, D\}, \Gamma_J \in \Theta}$  be sets of random variables such that

- for all  $t > 0$  and  $\Gamma_B \in \Theta$ ,  $B_{t-1,\Gamma_B}$  is  $\mathcal{A}_{\Delta t \cdot t}$ -measurable and has distribution  $\Gamma_B$ ,
- for all  $t > 0$ ,  $d \in \{1, \dots, D\}$  and  $\Gamma_J \in \Theta$ ,  $J_{t-1,d,\Gamma_J}$  is  $\mathcal{A}_{\Delta t \cdot t}$ -measurable and has distribution  $\Gamma_J$ ,
- for all  $t \geq 0$  and  $\Gamma_B, \Gamma_J \in \Theta$ , the random variables  $J_{t,1,\Gamma_J}, \dots, J_{t,D,\Gamma_J}, B_{t,\Gamma_B}$  are independent and
- for all  $t \geq 0$  and  $\Gamma_B, \Gamma_J \in \Theta$ ,  $\sigma(B_{t,\Gamma_B}, (J_{t,d,\Gamma_J})_{1 \leq d \leq D})$  is independent from  $\mathcal{A}_{\Delta t \cdot t}$ .

The  $\sigma$  operator provides the smallest  $\sigma$ -algebra, such that the arguments are measurable. A detailed introduction into  $\sigma$ -algebras can be found in [1]. Only a small portion of the introduced variables are needed, but if this huge set of random variables is not defined, random variables would have been used which exist only under certain conditions, which is formally not possible.

From now on the reality is approximated with the following model.

**Assumption 1** (Separation of logarithmic potential). *It is assumed that for objective functions  $f$  there exist  $\Delta t$ ,  $\Phi$ ,  $N_0$ ,  $c_0$  and  $c_s$  such that for any fixed time  $T_0 \cdot \Delta t$  there exist  $\mathcal{A}_{\Delta t \cdot T_0}$ -measurable  $\Theta$ -valued random variables  $\Gamma_{J,T_0}$  and  $\Gamma_{B,T_0}$  such that*

- for all  $T \geq T_0$  and  $d \in \{1, \dots, D\}$  the random variables  $B_{T,\Gamma_{B,T_0}}$  and  $J_{T,d,\Gamma_{J,T_0}}$  are  $\mathcal{A}_{(T+1)\Delta t}$ -measurable,

- for fixed times  $T_0\Delta t$  and  $T\Delta t$  and fixed dimension  $d$ ,  $P(A \setminus B) = 0$ , where

$$A := \{(\exists i : \alpha_i = T_0\Delta t \wedge \beta_i > T\Delta t) \wedge \Psi(T_0, d) \leq c_0 \wedge \max_{T_0 \leq t' \leq T} \Psi(t', d) \leq c_s\}$$

$$\text{and} \quad B := \{I_{T,d} = B_{T,\Gamma_{B,T_0}} + J_{T,d,\Gamma_{J,T_0}}\},$$

- for all  $\omega \in \Omega$  the expectation of a  $\Gamma_{J,T_0}(\omega)$ -distributed random variable is zero and
- for all  $\omega \in \Omega$  the first six moments of a random variable, which has distribution  $\Gamma_{B,T_0}(\omega)$  or  $\Gamma_{J,T_0}(\omega)$ , exist.

If there is no start of a stagnation phase at time  $T_0$  then  $\Gamma_{J,T_0}$  and  $\Gamma_{B,T_0}$  can have any value which does not contradict with the restrictions in Assumption 1. The value of  $\Gamma_{J,T_0}$  and  $\Gamma_{B,T_0}$  in these cases is not used in the analysis.

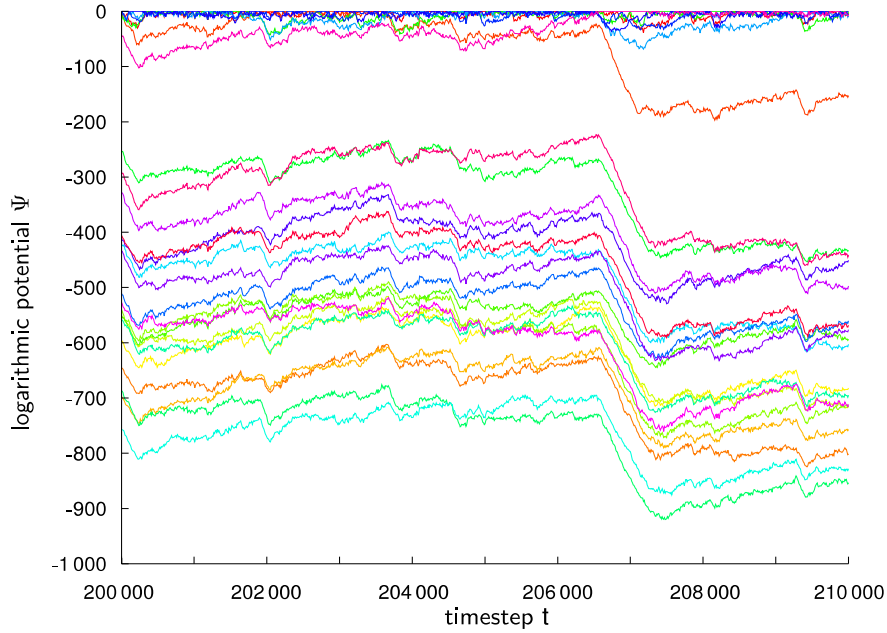


Figure 2: Logarithmic potential (with potential as specified in Definition 3) on sphere function in 25 dimensions and with 3 particles; each line represents the logarithmic potential of one dimension; visualized are time steps 200 000 to 210 000

The event  $A$  specifies that there is a stagnation phase which has started at time  $T_0\Delta t$  and has not ended at time  $T\Delta t$  or earlier and the dimension  $d$  is one of the dimensions with low logarithmic potential. This event implies almost surely the event  $B$  which states that the increments belonging to  $\Delta t$  time steps after time  $T\Delta t$  are composed by two components.

Mainly this assumption states that during stagnation phases the change in logarithmic potential of dimensions with low logarithmic potential is composed of some basic random noise ( $B_{T,\Gamma_{B,T_0}}$ ), which depends on the behavior of dimensions with high logarithmic potential, and some individual random noise ( $J_{T,d,\Gamma_{J,T_0}}$ ), which depends on the randomness in this specific dimension. The basic random noise ( $B_{T,\Gamma_{B,T_0}}$ ) will be called *base increments* and the individual random noise ( $J_{T,d,\Gamma_{J,T_0}}$ ) will be called *dimension dependent increments*. These two parts are independent of each other

and they are also independent for different times  $t$  as specified in Definition 5. The reason for this separation is that the dimensions with low logarithmic potential have no effect on local and global attractors of the swarm. Therefore these dimensions do not interact with each other. The distributions of the random variables  $B_{T,\Gamma_{B,T_0}}$  and  $J_{T,d,\Gamma_{J,T_0}}$  are fixed at time  $T_0$ , because  $\Gamma_{B,T_0}$  and  $\Gamma_{J,T_0}$  are  $\mathcal{A}_{T_0\Delta t}$ -measurable. The reason for this assumption is that according to the experiments the distributions mainly depend on the set of dimensions with low logarithmic potential, which is fixed at the beginning of a stagnation phase. Therefore  $\Theta$ , the set of probability distributions on  $\mathbb{R}$  specified in Definition 5, need to contain only very few distributions. The dimensions with high logarithmic potential determine the behavior of the particles, which partially determines the logarithmic potential of the dimensions with low logarithmic potential. An additional part of the logarithmic potential in a dimension arises from the random variables in the movement equations of the PSO in that dimension. These two parts determine the logarithmic potential of dimensions with low logarithmic potential. In Figure 2 you can see the similar behavior of dimensions with low logarithmic potential. Admittedly in any precise choice of a potential these new random variables will not be completely independent of each other, but if  $\Delta t$  is enlarged it can be expected that the dependencies become small. For example, if  $\Delta t$  is set to 1 then the increments describe the behavior of consecutive steps which depend on each other in a strong manner but for a larger  $\Delta t$  larger intervals depend only partially on each other. Therefore it is reasonable to accept the property of independence in the variable  $t$  for the theoretical model as an approximation of the real-world process.

Furthermore, the dependencies and independencies in Assumption 1 are not needed completely, but they are used to get some inequalities. These inequalities will be described and justified experimentally in Section 4.

**Definition 6** ( $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -Stagnation phase). *Let  $\mu \in \mathbb{R}$ ,  $M \in \mathbb{R}$ ,  $p_0 \in [0, 1]$ ,  $T_0 \geq 0$  be constants.*

*Let  $A$  be the event that a  $(\Delta t, \Phi, N_0, c_0, c_s)$ -stagnation phase starts at time  $T_0\Delta t$ , i. e.,*

$$A = \{\exists i \in \mathbb{N}_0 : \alpha_i = T_0\Delta t\}.$$

*Let  $C_1$  be the event that the expectation of a  $\Gamma_{B,T_0}$  distributed random variable is less or equal than  $\mu$ , i. e.,*

$$C_1 = \{E[B_{T_0,\Gamma_{B,T_0}} \mid \mathcal{A}_{T_0\Delta t}] \leq \mu\}.$$

*Let  $C_2$  be the event that the sixth moment of a  $\Gamma_{B,T_0}$  distributed random variable and a  $\Gamma_{J,T_0}$  distributed random variable is bounded by  $M$ , i. e.,  $H := E[B_{T_0,\Gamma_{B,T_0}} \mid \mathcal{A}_{T_0\Delta t}]$*

$$C_2 = \left\{ E \left[ (B_{T_0,\Gamma_{B,T_0}} - H + J_{t,1,\Gamma_{J,T_0}})^6 \mid \mathcal{A}_{T_0\Delta t} \right] \leq M \right\}.$$

*Let  $C_3$  be the event that a  $\Gamma_{B,T_0}$ -distributed random variable stays below  $\frac{\mu}{2}$  with a probability of at least  $p_0$ , i. e.,*

$$C_3 = \left\{ P \left( B_{T_0,\Gamma_{B,T_0}} - \frac{\mu}{2} \leq 0 \mid \mathcal{A}_{T_0\Delta t} \right) \geq p_0 \right\} = \left\{ E \left[ \mathbb{1}_{B_{T_0,\Gamma_{B,T_0}} - \frac{\mu}{2} \leq 0} \mid \mathcal{A}_{T_0\Delta t} \right] \geq p_0 \right\}.$$

*Let  $C_4$  be the event that a  $\Gamma_{J,T_0}$ -distributed random variable stays below  $-\frac{\mu}{2}$  with a probability of at least  $p_0$ , i. e.,*

$$C_4 = \left\{ P \left( J_{T_0,1,\Gamma_{B,T_0}} + \frac{\mu}{2} \leq 0 \mid \mathcal{A}_{T_0\Delta t} \right) \geq p_0 \right\} = \left\{ E \left[ \mathbb{1}_{J_{T_0,1,\Gamma_{B,T_0}} + \frac{\mu}{2} \leq 0} \mid \mathcal{A}_{T_0\Delta t} \right] \geq p_0 \right\}.$$



A  $(\Delta t, \Phi, N_0, c_0, c_s)$ -stagnation phase is called  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase if it achieves the conditions mentioned in  $C_1, C_2, C_3$  and  $C_4$ , i. e.,  $A \cap C_1 \cap C_2 \cap C_3 \cap C_4$  is called the event that a  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase starts at time  $T_0 \Delta t$ .

By definition the question whether a  $(\Delta t, \Phi, N_0, c_0, c_s)$ -stagnation phase, which starts at some time  $T_0 \cdot \Delta t$ , is a  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase can be answered at the beginning of that phase, i. e., the indicator function of this event is  $\mathcal{A}_{T_0 \cdot \Delta t}$ -measurable.

**Definition 7** (Good stagnation phase). A  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase is called good if  $\mu < 0$  and  $p_0 > 0$ .

These definitions only classify stagnation phases. For a single stagnation phase which has negative expectation and bounded sixth moment  $\mu$  and  $M$  can be chosen as the exact values. Furthermore, the probabilities which should be bounded below by  $p_0$  are positive because the probability that a random variable encounters a value less or equal its expectation or even larger values is always positive. Such stagnation phases are called good, because it can be proved that those phases do not end with positive probability.

The following theorem shows that, with positive probability, the swarm never recovers from encountering a good stagnation phase.

**Theorem 1.** A good  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase does not end with a probability, which is at least  $p := p(N_0, \mu, M, p_0) > 0$ .

*Proof.* Let  $X_0, V_0$  and  $L_0$  be initialized such that a good  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase starts at time 0 and let  $\Gamma_J$  and  $\Gamma_B$  be the related probability distributions of this stagnation phase specified in Assumption 1. As the PSO-process is a Markov-process, evaluating the process which starts in a good  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase is equivalent to evaluate a good  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase within a PSO run. Let  $S_0$  be  $\{d \in \{1, \dots, D\} \mid \Psi(0, d) \leq c_0\}$ , the starting subset.  $S_0, \Gamma_B$  and  $\Gamma_J$  are fixed objects here, because they are always known at the beginning of a stagnation phase. Then the probability that this good stagnation phase does not end is equal to

$$\mathbb{P}\left(\exists U \subset S_0 : |U| \geq N_0 \wedge \sup_{d \in U, t > 0} \Psi(t, d) \leq c_s\right).$$

Let  $S$  be defined as  $\{d \in S_0 : |\{1, \dots, d\} \cap S_0| \leq N_0\}$ , the subset of  $S_0$ , which contains the  $N_0$  smallest indices.  $S$  is a fixed object too. The probability that this good stagnation phase does not end is at least  $\mathbb{P}(\sup_{d \in S, t > 0} \Psi(t, d) \leq c_s)$ . Let  $\tilde{I}_{t,d} := B_{t,\Gamma_B} + J_{t,d,\Gamma_J}$  (see Assumption 1), then  $I_{t,d} \equiv \tilde{I}_{t,d}$  if the stagnation phase has not ended at time  $t$  or earlier and  $d$  is in the set  $\{\tilde{d} \in S_0 \mid \max_{0 < \tilde{t} \leq t} \Psi(\tilde{t}, \tilde{d}) \leq c_s\}$ . Therefore  $I_{t,d} \equiv \tilde{I}_{t,d}$  for all  $d \in S$  if  $\max_{\tilde{d} \in S, 0 < \tilde{t} \leq t} \Psi(\tilde{t}, \tilde{d}) \leq c_s$ . It follows

$$\begin{aligned} & \mathbb{P}\left(\sup_{t > 0, d \in S} \Psi(t, d) \leq c_s\right) \\ &= \mathbb{P}\left(\sup_{t > 0, d \in S} \sum_{\tilde{t}=0}^{t-1} \tilde{I}_{\tilde{t},d} + \Psi(0, d) \leq c_s\right) \\ &\geq \mathbb{P}\left(\sup_{t > 0, d \in S} \sum_{\tilde{t}=0}^{t-1} \tilde{I}_{\tilde{t},d} + c_0 \leq c_s\right) \end{aligned}$$

$$\begin{aligned}
&= \mathbb{P}\left(\sup_{t>0, d \in S} \sum_{\tilde{t}=0}^{t-1} \tilde{I}_{\tilde{t},d} \leq c_s - c_0\right) \\
&\geq \mathbb{P}\left(\sup_{t>0, d \in S} \sum_{\tilde{t}=0}^{t-1} \tilde{I}_{\tilde{t},d} \leq 0\right) \\
&= \mathbb{P}\left(\sup_{t>0, d \in S} \sum_{\tilde{t}=0}^{t-1} \left(\left(B_{\tilde{t},\Gamma_B} - \frac{\mu}{2}\right) + \left(J_{\tilde{t},d,\Gamma_J} + \frac{\mu}{2}\right)\right) \leq 0\right) \\
&\geq \mathbb{P}\left(\sup_{t>0} \sum_{\tilde{t}=0}^{t-1} \left(B_{\tilde{t},\Gamma_B} - \frac{\mu}{2}\right) \leq 0 \wedge \sup_{t>0, d \in S} \sum_{\tilde{t}=0}^{t-1} \left(J_{\tilde{t},d,\Gamma_J} + \frac{\mu}{2}\right) \leq 0\right) \\
&= \mathbb{P}\left(\sup_{t>0} \sum_{\tilde{t}=0}^{t-1} \left(B_{\tilde{t},\Gamma_B} - \frac{\mu}{2}\right) \leq 0\right) \prod_{d \in S} \mathbb{P}\left(\sup_{t>0} \sum_{\tilde{t}=0}^{t-1} \left(J_{\tilde{t},d,\Gamma_J} + \frac{\mu}{2}\right) \leq 0\right).
\end{aligned}$$

The expectations of  $\left(B_{\tilde{t},\Gamma_B} - \mu/2\right)$  and  $\left(J_{\tilde{t},d,\Gamma_J} + \mu/2\right)$  are at most  $\mu/2$ , which is less than zero. With Lemma 3 the first six moments of

$$\left(B_{\tilde{t},\Gamma_B} - \mu/2 - \mathbb{E}[B_{\tilde{t},\Gamma_B} - \mu/2]\right) \text{ and } \left(J_{\tilde{t},d,\Gamma_J} + \mu/2 - \mathbb{E}[J_{\tilde{t},d,\Gamma_J} + \mu/2]\right)$$

can be bounded by  $h(M)$ . Therefore Lemma 2 tells us that

$$\mathbb{P}\left(\sup_{t>0} \sum_{\tilde{t}=0}^{t-1} (B_{\tilde{t},\Gamma_B} - \mu/2) \leq 0\right) \geq p(\mu/2, h(M), p_0) > 0$$

and

$$\mathbb{P}\left(\sup_{t>0} \sum_{\tilde{t}=0}^{t-1} (J_{\tilde{t},d,\Gamma_J} + \mu/2) \leq 0\right) \geq p(\mu/2, h(M), p_0) > 0.$$

Altogether we have

$$\begin{aligned}
&\mathbb{P}(\exists U \subset S_0 : |U| \geq N_0 \wedge \sup_{d \in U, t>0} \Psi(t, d) \leq c_s) \\
&\geq \mathbb{P}\left(\sup_{t>0, d \in S} \Psi(t, d) \leq c_s\right) \\
&\geq \mathbb{P}\left(\sup_{t>0} \sum_{\tilde{t}=0}^{t-1} \left(B_{\tilde{t},\Gamma_B} - \frac{\mu}{2}\right) \leq 0\right) \prod_{d \in S} \mathbb{P}\left(\sup_{t>0} \sum_{\tilde{t}=0}^{t-1} \left(J_{\tilde{t},d,\Gamma_J} + \frac{\mu}{2}\right) \leq 0\right) \\
&\geq p\left(\frac{\mu}{2}, h(M), p_0\right) \cdot p\left(\frac{\mu}{2}, h(M), p_0\right)^{N_0} \\
&=: p(N_0, \mu, M, p_0) =: p
\end{aligned}$$

□

According to Donsker's theorem [2] incremental processes composed of independent identical distributed increments converge to a Brownian Motion if they are correctly scaled. As our general model uses independent increments, we also discuss the Brownian Motion as a feasible model for our

process. To acquire Brownian Motion it is additionally assumed in the model that the increments are normally distributed during good stagnation phases. Then a Brownian Motion, with some variance  $\sigma^2$  and negative drift  $\mu$  according to the expectation and variance of the increments, approximates the logarithmic potential. I. e., the Brownian Motion at discrete times would be equal to the logarithmic potential at these times. In [14], it is proved that the probability that a standard Brownian Motion crosses some line  $y = m \cdot x + t$  is equal to  $\exp(-2 \cdot m \cdot t)$  for positive values  $m$  and  $t$ . Under these assumptions,  $\left(\sum_{\tilde{t}=T_0}^{t-1} (\tilde{I}_{\tilde{t},d} - \mu)\right) / \sigma$  can be approximated by a standard Brownian Motion, i. e., a Brownian Motion with expectation 0 and variance 1 per time step. The previous bound  $c_s$  is now moved to  $-\frac{\mu}{\sigma}t + \frac{c_s - c_0}{\sigma}$ . If the Brownian Motion stays below this line, then it also stays below the line at the discrete time steps referring to time steps of the PSO, which implies that the logarithmic potential stays below  $c_s$  indefinitely. Hence, the probability that the logarithmic potential stays below  $c_s$  forever, is approximated by  $1 - \exp(2 \frac{(c_s - c_0)\mu}{\sigma^2})$ , where  $\mu$  and  $\sigma^2$  represent the expectation and variance of the increments per  $\Delta t$  time steps. An approximate lower bound for the probability that at least  $N_0$  dimensions in  $S_0$  stay below  $c_s$  is represented by the  $N_0$ 'th power of the probability in one dimension:  $(1 - \exp(2 \frac{(c_s - c_0)\mu}{\sigma^2}))^{N_0}$ . In the previous model we were only able to determine some positive lower bound for the probability that stagnation remains indefinitely. This model supplies the previously mentioned formula, which approximates the actual probability quite well for at least  $N_0 = 1$ . In Section 4.5 the values obtained from that formula will be compared to experimentally measured values.

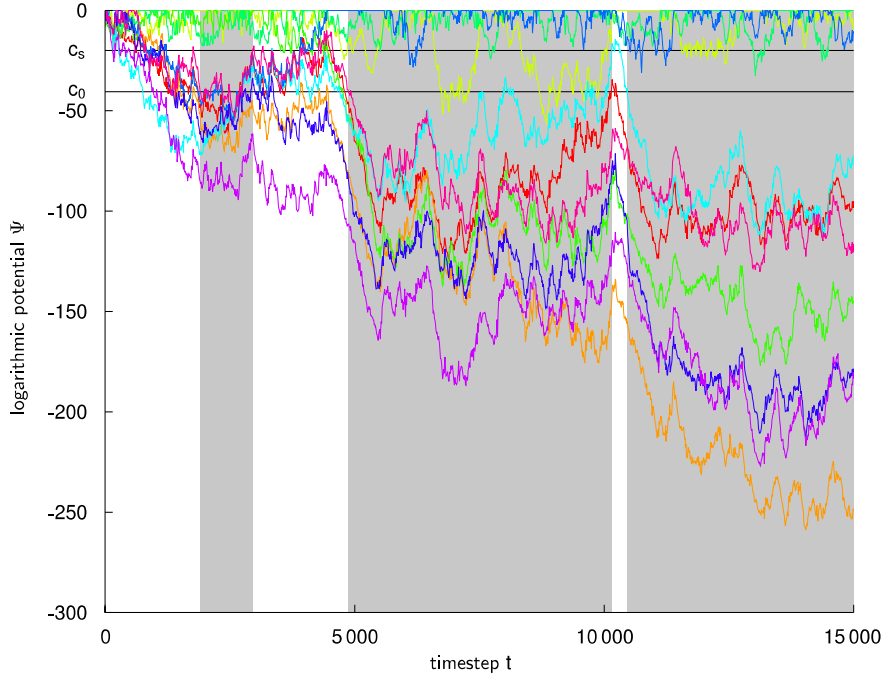


Figure 3: Logarithmic potential (with potential as specified in Definition 3) on sphere function in 10 dimensions and with 2 particles; each line represents the logarithmic potential of one dimension,  $c_0 = -40$ ,  $c_s = -20$ ,  $N_0 = 7$

### 3.4 Convergence to Non-Optimal Points

**Definition 8** (Phases). Let  $\Delta t, \Phi, N_0, c_0, c_s, \mu, M$  and  $p_0$  be fixed objects. Let  $(\alpha_i)_{i \in \mathbb{N}_0}$  and  $(\beta_i)_{i \in \mathbb{N}_0}$  be the stopping times specified in Definition 4. Furthermore, the following stopping times are defined:

$$\tilde{\beta}_{-1} = 0$$

and inductively for all  $i \geq 0$ :

$$\begin{aligned}\tilde{\alpha}_i &= \inf\{t \geq \tilde{\beta}_{i-1} : \exists j \in \mathbb{N}_0 : \alpha_j = t \wedge \text{the } (\Delta t, \Phi, N_0, c_0, c_s)\text{-stagnation phase} \\ &\quad \text{starting at time } t \text{ is a } (\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)\text{-stagnation phase}\}, \\ \tilde{\beta}_i &= \inf\{t > \tilde{\alpha}_i : \exists j \in \mathbb{N}_0 : \beta_j = t\}.\end{aligned}$$

$\tilde{\alpha}_i$  and  $\tilde{\beta}_i$  determine the start and end of the  $i$ 'th  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase. Phases are random time intervals. Phases of type  $PH_X$  are phases before or between  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phases. The  $i$ 'th phase of type  $PH_X$  is defined as

$$\begin{aligned}&[\tilde{\beta}_{i-1}, \tilde{\alpha}_i] && \text{if } \tilde{\alpha}_i < \infty \\ &[\tilde{\beta}_{i-1}, \infty[ && \text{if } \tilde{\alpha}_i = \infty \wedge \tilde{\beta}_{i-1} < \infty \\ &\emptyset && \text{otherwise.}\end{aligned}$$

Phases of type  $PH_Y$  are  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phases which have an end. The  $i$ 'th phase of type  $PH_Y$  is defined as

$$\begin{aligned}&[\tilde{\alpha}_i, \tilde{\beta}_i] && \text{if } \tilde{\beta}_i < \infty \\ &\emptyset && \text{otherwise.}\end{aligned}$$

The phase of type  $PH_F$  is a  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase which has no end and it is defined as

$$\begin{aligned}&[\tilde{\alpha}_i, \infty[ && \text{if } \tilde{\alpha}_i < \infty \wedge \tilde{\beta}_i = \infty \\ &\emptyset && \text{if no such } i \text{ exists.}\end{aligned}$$

$X_i$  is defined to be the duration of the  $i$ 'th phase of type  $PH_X$ :

$$X_i := \begin{cases} \tilde{\alpha}_i - \tilde{\beta}_{i-1} & \text{if } \tilde{\beta}_{i-1} < \infty \\ 0 & \text{otherwise.} \end{cases}$$

$Y_i$  is defined to be the duration of the  $i$ 'th phase of type  $PH_Y$ :

$$Y_i := \begin{cases} \tilde{\beta}_i - \tilde{\alpha}_i & \text{if } \tilde{\beta}_i < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, the random variable  $T_F$  is defined as the time when the final phase of type  $PH_F$  starts:

$$T_F := \sum_{i=0}^{\infty} X_i + Y_i.$$

$Y_i$  is a  $\mathbb{N}_0$ -valued random variable and  $X_i, T_F$  are  $\mathbb{N}_0 \cup \{\infty\}$ -valued random variables.

A phase of type  $PH_X$  is the interval beginning at the start of the PSO or at the end of a  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase and ending at the start of the next  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase. If there is no further  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase then this phase does not end. A phase of type  $PH_Y$  is the time interval of a  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase which has an end and a phase of type  $PH_F$  is the time interval of a  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase which has no end. The duration of a phase equals the difference of the last time step and the first time step contained in the related time interval and is infinity if the time interval has no limit. If the  $i$ 'th instance of type  $PH_X$ -phase of a specific run does not exist then  $X_i$  is zero. If the  $i$ 'th instance of type  $PH_Y$ -phase of a specific run does not exist then  $Y_i$  is zero.  $T_F$  is infinity if no phase of type  $PH_F$  starts. The process starts with a phase of type  $PH_X$ . Phases of type  $PH_X$  and  $PH_Y$  are alternating until the final phase, which has type  $PH_F$ , begins. Alternatively the last phase could be a phase of type  $PH_X$  or phases of type  $PH_X$  and  $PH_Y$  are alternating infinitely often, but it will be shown in Theorem 2 that this does not happen almost surely. It may occur that a phase of type  $PH_X$  has duration zero if the set of dimensions with low logarithmic potential has changed. Figure 3 shows an example run of a 10-dimensional PSO with 2 particles.  $c_0 := -40$ ,  $c_s := -20$  and  $N_0 := 7$ , i. e., a stagnation phase starts whenever 7 or more of the 10 dimensions have logarithmic potential below  $-40$ .

In the following definition a group of functions is specified, which allows for comprehensible assumptions.

**Definition 9** (Composite functions). *A function is called composite if it can be written as  $f(x) = g(\sum_{i=1}^D f_i(x_i))$ , where  $g$  is a strictly monotonically increasing function and  $(f_i)_{i \in \{1, \dots, D\}}$  are functions which have a lower bound each.*

For example all  $q$ -norms  $(\sum_{i=1}^D |x_i|^q)^{\frac{1}{q}}$  are composite functions with  $f_i(x_i) := |x_i|^q$  and  $g(y) := \text{sign}(y)|y|^{\frac{1}{q}}$ , where  $\text{sign}$  represents the signum function  $\text{sign}(y) := 1$  if  $y > 0$ , 0 if  $y = 0$  and  $-1$  if  $y < 0$ .

Furthermore, the following notation will be used:

**Definition 10.** *For objects  $W = (W^1, \dots, W^D)$  and  $S \subset \{1, \dots, D\}$  we write  $W^{d \in S}$  for the object, where all entries in dimensions contained in  $S$  remain, i. e., let  $d_1$  to  $d_{|S|}$  be the dimensions contained in  $S$  in increasing order then  $W^{d \in S} = (W^{d_1}, \dots, W^{d_{|S|}})$ . Analogously  $W^{d \notin S}$  is written for  $W^{d \in \{1, \dots, D\} \setminus S}$ , where all entries in dimensions contained in  $S$  are removed.*

**Assumption 2** (Insignificance of stagnating dimensions (1)). *Let  $T_0 \in \mathbb{N}$ ,  $T \in \mathbb{N}$  and let  $S$  be a subset of  $\{1, \dots, D\}$ , such that  $T_0 < T$  and  $|S| \geq N_0$ . For fixed objects  $\Delta t, \Phi, N_0, c_0, c_s$  and a composite objective function  $f(x) = g(\sum_{i=1}^D f_i(x_i))$  a modified PSO can be defined. The modified PSO starts at time  $T_0 \Delta t$  with initialization*

$$\tilde{X}_{T_0 \Delta t} := X_{T_0 \Delta t}^{d \notin S}, \tilde{V}_{T_0 \Delta t} := V_{T_0 \Delta t}^{d \notin S}, \tilde{L}_{T_0 \Delta t} := L_{T_0 \Delta t}^{d \notin S} \text{ and}$$

$$\text{objective function } \tilde{f} := \sum_{i \notin S} f_i(x_i).$$

Furthermore, similar to Definition 1 the movement equations for the modified PSO for all particles

$n$  ( $1 \leq n \leq N$ ) are defined to be

$$\begin{aligned}\tilde{G}_t^n &:= \operatorname{argmin}_{x \in \bigcup_{i=1}^{n-1} \{\tilde{L}_{t+1}^i\} \cup \bigcup_{i=n}^N \{\tilde{L}_t^i\}} \tilde{f}(x) \text{ for } t \geq T_0 \Delta t, \\ \tilde{V}_{t+1}^n &:= \chi \cdot \tilde{V}_t^n + c_1 \cdot (r_t^n)^{d \notin S} \odot (\tilde{L}_t^n - \tilde{X}_t^n) + c_2 \cdot (s_t^n)^{d \notin S} \odot (\tilde{G}_t^n - \tilde{X}_t^n) \text{ for } t \geq T_0 \Delta t, \\ \tilde{X}_{t+1}^n &:= \tilde{X}_t^n + \tilde{V}_{t+1}^n \text{ for } t \geq T_0 \Delta t, \\ \tilde{L}_{t+1}^n &:= \operatorname{argmin}_{x \in \{\tilde{X}_{t+1}^n, \tilde{L}_t^n\}} \tilde{f}(x) \text{ for } t \geq T_0 \Delta t.\end{aligned}$$

with reused random variables  $(r_t^n)^{d \notin S}$  and  $(s_t^n)^{d \notin S}$ . Let  $A$  and  $B$  be the events

$$A := \{\exists i \in \mathbb{N}_0 : \alpha_i = T_0 \Delta t \wedge \max_{d \in S} \Psi(T_0, d) \leq c_0 \wedge \max_{d \in S, T_0 \leq t \leq T} \Psi(t, d) \leq c_s\}$$

and

$$B := \{X_t^{d \notin S} = \tilde{X}_t \wedge V_t^{d \notin S} = \tilde{V}_t \wedge L_t^{d \notin S} = \tilde{L}_t \forall t \in \{T_0 \Delta t, \dots, T \Delta t - 1\}\}.$$

It is assumed that there exist  $\Delta t$ ,  $\Phi$ ,  $N_0$ ,  $c_0$  and  $c_s$  such that Assumption 1 is applicable and  $P(A \setminus B) = 0$ .

The modified PSO represents the actual PSO, but the dimensions with low logarithmic potential are completely removed. The event  $A$  appears if a stagnation phase starts at time  $T_0 \Delta t$  and has not ended at time  $T \Delta t$  and  $S$  is the set of stagnating dimensions. The event  $B$  appears if the reduced process and the actual process are equal at times  $t$  such that  $T_0 \leq t < T$  for all dimensions not in  $S$ . The assumption states that almost surely  $A$  implies  $B$ , which means that this modified PSO behaves similar to the actual PSO if a stagnation phase starts at time  $T_0$  and all dimensions in  $S$  are stagnating. Therefore stagnating dimensions have no effect on the other dimensions, but dimensions not in  $S$  have effect on the stagnating dimensions because the times when attractors are updated are determined by dimensions not in  $S$ .

In the following a short motivation is presented why this assumption is reasonable. The potential used in this paper represents the impact of a single dimension to the change in the function value in a single step. The probability that a dimension can change the evaluation whether a new position is a local or global attractor can therefore approximately be bounded by  $2^{\Psi(t,d)}$ . For stagnation phases with  $c_s = -20$  this value is approximately 0.000 001 and therefore the expected waiting time that a dimension with that low potential has effect on the behavior of the swarm is approximately bounded below by 1 000 000 iterations. Furthermore, for good stagnation phases the logarithmic potential is expected to decrease linearly on average. Neglecting the randomness the logarithmic potential at time  $t$  can be bounded by  $c_0 - \mu(t - T_0 \Delta t)$ , where  $-\mu$  is the negative expectation of the potential per time step, and therefore the expected number of time steps such that this dimension has effect on a decision whether a new position is a local or global attractor after the beginning of an unlimited stagnation phase is approximately bounded by

$$\sum_{i>0} 2^{c_0 - \mu i} \leq 2^{c_0} \int_0^\infty \exp(-\ln(2)\mu i) di = 2^{c_0} \frac{1}{\ln(2)\mu},$$

which is a finite value. Furthermore,  $c_0$  can be chosen such that this value is even less than one. Also changes in dimensions with low logarithmic potential can not accumulate over time, because

if updates in local and global attractors happen quite often, then accumulation can not appear and if the updates occur not quite often, then changes will not accumulate as well because the positions of the particles will mainly stay between the local and the global attractors.

The last assumption states that dimensions, which have low logarithmic potential, have no effect on the swarm. This means that those dimensions receive a series of decision results whether a position is a new local or global attractor. Those decision results influence how the positions in dimensions with low logarithmic potential develop.

In this paper the behavior of the PSO for large time scales is analyzed. Therefore it is necessary to define in which cases convergence is present.

**Definition 11** (Swarm convergence). *A swarm is said to converge if the coordinates of the global and local attractors and the positions converge to the same point and the velocities of the particles converge to zero. The limit of the positions is also called point of convergence.*

If the velocity tends exponentially to zero then convergence of attractors and positions also appears.

If there is a stagnation phase which has no end then there is a non empty subset of dimensions and the dimensions in this set have low logarithmic potential as soon as the last stagnation phase starts. Assumption 2 states that there exists an alternative PSO without that dimensions. Although these dimensions have no effect on the remaining dimensions, these dimensions are still changing. The dimensions with low logarithmic potential receive a series of updates of the local and global attractors from the reduced process. If convergence appears, some additional conclusions can be made. Depending on the update series and the random variables for the dimensions with low logarithmic potential, some final value for that dimensions is reached. As it is not plausible that the limit of the coordinates for that dimensions is fixed at the start of the last stagnation phase, it is assumed that there is no limit value of these coordinates, which has positive probability. Neither optimal values nor any other value.

Therefore the following assumption is concluded.

**Assumption 3** (Insignificance of stagnating dimensions (2)). *For fixed objects  $\Delta t$ ,  $\Phi$ ,  $N_0$ ,  $c_0$ ,  $c_s$ ,  $\mu$ ,  $M$  and  $p_0$  let  $X_0$ ,  $V_0$  and  $L_0$  be initialized such that it represents a possible start of a good  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase, let  $\Gamma_J$  and  $\Gamma_B$  be the related probability distributions of this stagnation phase specified in Assumption 1 and let*

$$S \subset \{d \in \{1, \dots, D\} \mid \Psi(0, d) \leq c_0\}$$

*such that  $|S| = N_0$ . Let  $\mathcal{E}$  be the event*

$$\mathcal{E} := \left\{ \sup_{t \geq 0, d \in S} \Psi(t, d) \leq c_s \wedge \text{the swarm converges} \right\}.$$

*It is assumed that there exist  $\Delta t$ ,  $\Phi$ ,  $N_0$ ,  $c_0$ ,  $c_s$ ,  $\mu$ ,  $M$  and  $p_0$  such that Assumption 1 is applicable and*

$$P\left(\lim_{t \rightarrow \infty} X_t^{n,d} = v \mid \mathcal{E}\right) = 0 \text{ for all } n \in \{1, \dots, N\}, v \in \mathbb{R} \text{ and } d \in S.$$

This assumption states that if the PSO reaches a stagnation phase which does not end and the PSO converges, then it is almost surely true that specific points will not be reached, because at least the coordinates of the insignificant dimensions will differ almost surely. This implies that if

the set of local optima is countable then almost surely none of the local optima will be reached, because the set of possible coordinate values of optimal points is countable for each dimension and therefore the dimensions with low logarithmic potential will not reach any of those values almost surely. Therefore stagnation occurs in those cases.

Assumption 3 is not restricted to composite functions. The behavior for functions which are not composite is more complicated, but the effect is the same. An aspect which will not be proved is that dimensions which become relevant from time to time are optimizing. If the coordinates in dimensions with low logarithmic potential are converging to a non-optimal value then the optimal value for the coordinates, which are optimizing, is probably not the same as in the case where all dimensions are optimizing. For example the function  $f(x, y) = 2(x+y)^2 + (x-y)^2 = 3x^2 + 2xy + 3y^2$  is minimal for  $x = 0$  and  $y = 0$ , but if  $x$  tends to 1 then  $f(1, y) = 3y^2 + 2y + 3$  is minimal for  $y = -\frac{1}{3} \neq 0$ . Nevertheless, the effect is the same for functions, which are not composite, because the influence of the dimensions with low logarithmic potential is heavily delayed. The scales of the change in stagnating dimensions and non-stagnating dimensions are very different and therefore the change in the optimal position is recognized by the optimizing dimensions later in the process, when the stagnating dimensions do not have the possibility to change the optimal position and their position in a comparable manner.

For example if the logarithmic potential of a dimension  $d$  stays smaller than  $-100$  then (for the specified experimental potential and tested functions) also the velocities in dimension  $d$  are by a factor of  $2^{-100}$  smaller than the velocities in the dimensions with most potential. If an attractor is updated, then its  $d$ -th position entry changes a bit. This change might result in a minor change of the optimal position for all other dimensions if the position value in dimension  $d$  is regarded as a constant value, i.e. let  $f_c(x_1, \dots, x_{d-1}, x_{d+1}, \dots, x_D) := f(x_1, \dots, x_D)$ , let  $c_{old}$  be the previous value of the  $d$ -th position entry and let  $c_{new}$  be the new value of the  $d$ -th position entry, then the optimal position of  $f_{c_{old}}$  and  $f_{c_{new}}$  may vary a bit. Currently the attractor choice does hardly depend on that change, but if the swarm finally converges, then the velocity in dimensions with most logarithmic potential finally reaches the same scale as the previous change of the optimal position where the position value in dimension  $d$  is regarded as a constant. The velocity in dimension  $d$  has also decreased and is again much smaller. Therefore it cannot undo the changes done many steps in the past, where the velocity in dimension  $d$  had the same scale as the dimensions with large logarithmic potential now have. This is meant as the heavily delayed influence of dimensions with low logarithmic potential. As the movement in dimension  $d$  was at no time guided to the optimal point in this dimension, it is almost impossible that this dimension is optimizing.

With Assumption 3 the main theorem of this paper can be proved.

**Theorem 2.** *Let  $\Phi$  be a potential,  $c_0 \leq c_s < 0$ ,  $\Delta t > 0$ ,  $N_0 > 0$  be constants such that the Assumptions 1 and 3 are applicable. Furthermore, let  $T > 0$ ,  $\mu < 0$ ,  $M < \infty$  and  $p_0 > 0$  be constants. Let  $A_{X,i} := \{\hat{\beta}_{i-1} < \infty\}$  for all  $i$ . If the following conditions hold:*

- *The swarm converges almost surely,*
- *the objective function  $f$  has only a countable number of local optima and*
- *for all  $i$  the expectation  $E[X_i \mid A_{X,i}] \leq T$  if  $P(A_{X,i}) > 0$ ,*

*then the swarm converges to a point, which is not a local optimum, almost surely.*



*Proof.* If the final phase which has type  $PH_F$  appears, then the PSO will not converge to a local optimum for reasons explained below Assumption 3. Now all which is needed to be shown is that a phase of type  $PH_F$  appears almost surely. This is true if it can be shown that the expected value of  $T_F$ , the starting time of the phase of type  $PH_F$ , is finite.

Let  $A_{X,i} := \{\tilde{\beta}_{i-1} < \infty\}$ , the event that the  $i$ 'th instance of a phase which has type  $PH_X$  appears, let  $A_{Y,i} := \{\tilde{\beta}_i < \infty\}$ , the event that the  $i$ 'th instance of a phase which has type  $PH_Y$  appears and let  $A_{S,i} := \{\tilde{\alpha}_i < \infty\}$ , the event that the  $i$ 'th instance of a good  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase appears. Let  $N_X$  be  $\sup\{i \in \mathbb{N} : P(A_{X,i}) > 0\}$  and let  $N_Y$  be  $\sup\{i \in \mathbb{N} : P(A_{Y,i}) > 0\}$ . Both values can be infinity. Let  $S_{0,i}$  be the random variable, which represents the starting subset of the  $i$ 'th good  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase or the empty set if this phase does not occur. Let  $\tilde{I}_{t,d,i}$  be  $B_{t,\tilde{\Gamma}_{B,i}} + J_{t,d,\tilde{\Gamma}_{J,i}}$ , similar as in the proof of Theorem 1, such that  $\tilde{\Gamma}_{B,i}$  and  $\tilde{\Gamma}_{J,i}$  are the random probability distributions of the  $i$ 'th good  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase. The random probability distributions are introduced in Assumption 1. If the  $i$ 'th good  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase does not exist then  $\tilde{\Gamma}_{B,i}$  is defined to be a Dirac impulse on  $\mu$  and  $\tilde{\Gamma}_{J,i}$  is defined to be a Dirac impulse on zero. Additionally, let  $\mathcal{A}$  be the generated  $\sigma$ -Algebra of the family of  $\sigma$ -Algebras  $\{\mathcal{A}_t\}_{t \in \mathbb{N}}$ , where  $\mathcal{A}_t$  is the natural filtration of the PSO as specified in Definition 1. This  $\sigma$ -algebra is needed because for any time  $t$  it is not possible to decide whether a stagnation phase, which is active at time  $t$  has type  $PH_F$  or  $PH_Y$ . Therefore  $A_{Y,i}$  is not necessarily contained in  $\mathcal{A}_t$  for any  $t$ . The conditional expectation of  $X_i$ , the duration of the  $i$ 'th phase of type  $PH_X$ , is by the requirements of this theorem bounded by

$$E[X_i \mid A_{X,i}] \leq T$$

if the event  $A_{X,i}$  has positive probability. Let  $T_{S,i}$  be the random variable, such that  $T_{S,i} \cdot \Delta t$  represents the time of the start of the  $i$ 'th stagnation phase, which has either type  $PH_Y$  or  $PH_F$ . If the  $i$ 'th stagnation phase does not occur then  $T_{S,i}$  is set to infinity. As already specified,  $Y_i$  represents the duration of the  $i$ 'th phase of type  $PH_Y$  if that phase exists and zero otherwise. Let  $p$  be defined as  $p(N_0, \mu, M, p_0)$ , the lower bound of the probability that a good  $(\Delta t, \Phi, N_0, c_0, c_s, \mu, M, p_0)$ -stagnation phase does not end from Theorem 1.

$$\begin{aligned} E[T_F] &= \sum_{i=0}^{N_X} P(A_{X,i}) E[X_i \mid A_{X,i}] + \sum_{i=0}^{N_Y} \underbrace{P(A_{Y,i}) E[Y_i \mid A_{Y,i}]}_{=E[Y_i]} \\ &= \sum_{i=0}^{N_X} P(A_{X,i}) E[X_i \mid A_{X,i}] + \sum_{i=0}^{N_Y} P(A_{S,i}) E[Y_i \mid A_{S,i}] \end{aligned}$$

compare with end of stagnation phase (Definition 4)

$$\begin{aligned} &= \sum_{i=0}^{N_X} P(A_{X,i}) E[X_i \mid A_{X,i}] + \sum_{i=0}^{N_Y} P(A_{S,i}) \sum_{t=1}^{\infty} t \Delta t \cdot \\ &\quad \cdot P(|\{d \in S_{0,i} : \sup_{0 \leq \tilde{t} < t} \Psi(T_{S,i} + \tilde{t}, d) \leq c_s\}| \geq N_0 \\ &\quad \wedge |\{d \in S_{0,i} : \sup_{0 \leq \tilde{t} \leq t} \Psi(T_{S,i} + \tilde{t}, d) \leq c_s\}| < N_0 \mid A_{S,i}) \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i=0}^{N_X} (1-p)^i \mathbb{E}[X_i \mid A_{X,i}] + \sum_{i=0}^{N_Y} (1-p)^i \sum_{t=1}^{\infty} t \Delta t \cdot \\
&\quad \cdot \mathbb{P}(|\{d \in S_{0,i} : \sup_{0 \leq \tilde{t} < t} \sum_{t'=T_{S,i}}^{T_{S,i}+\tilde{t}} \tilde{I}_{t',d,i} \leq 0\}| \geq N_0 \wedge |\{d \in S_{0,i} : \sup_{0 \leq \tilde{t} \leq t} \sum_{t'=T_{S,i}}^{T_{S,i}+\tilde{t}} \tilde{I}_{t',d,i} \leq 0\}| < N_0 | A_{S,i}) \\
&\leq \sum_{i=0}^{\infty} (1-p)^i T + \sum_{i=0}^{N_Y} (1-p)^i \sum_{t=1}^{\infty} t \Delta t \cdot \mathbb{P}(\exists d \in S_{0,i} : \sup_{0 \leq \tilde{t} < t} \sum_{t'=T_{S,i}}^{T_{S,i}+\tilde{t}} \tilde{I}_{t',d,i} \leq 0 \wedge \sum_{t'=T_{S,i}}^{T_{S,i}+t} \tilde{I}_{t',d,i} > 0 | A_{S,i}) \\
&\leq \frac{T}{p} + \sum_{i=0}^{N_Y} (1-p)^i \sum_{t=1}^{\infty} t \Delta t \sum_{d=1}^D \mathbb{P}(d \in S_{0,i} \mid A_{S,i}) \cdot \mathbb{P}(\sup_{0 \leq \tilde{t} < t} \sum_{t'=T_{S,i}}^{T_{S,i}+\tilde{t}} \tilde{I}_{t',d,i} \leq 0 \wedge \sum_{t'=T_{S,i}}^{T_{S,i}+t} \tilde{I}_{t',d,i} > 0 | A_{S,i}) \\
&\leq \frac{T}{p} + \sum_{i=0}^{N_Y} (1-p)^i \sum_{t=1}^{\infty} t \Delta t \sum_{d=1}^D \mathbb{P}(\sum_{t'=T_{S,i}}^{T_{S,i}+t} \tilde{I}_{t',d,i} > 0 | A_{S,i}) \\
&\stackrel{\text{Lemma 2}}{\leq} \frac{T}{p} + \sum_{i=0}^{N_Y} (1-p)^i \sum_{t=1}^{\infty} t \Delta t \sum_{d=1}^D \frac{C(\mu, h(M))}{t^3} \\
&\leq \underbrace{\frac{T}{p} + \frac{1}{p} D \Delta t \cdot C(\mu, h(M)) \sum_{t=1}^{\infty} t^{-2}}_{=: C(\Delta t, N_0, \mu, M, p_0, T) < \infty}.
\end{aligned}$$

□

This is the final statement of the theoretical part. The proof of Theorem 2 not only shows that the final phase with type  $PH_F$  appears almost surely, but it also verifies that the expected begin of the final phase is a finite value. Now evidence is provided in the experimental part, that those conditions truly occur, and therefore that PSO does not converge to a local optimum almost surely for some well known benchmarks and parameters of the PSO, which are commonly used.

## 4 Experimental Results

Since we are interested in the behavior of the swarm for  $t \rightarrow \infty$ , we let the swarm do a very large number of iterations. During the ongoing process, the absolute value of the velocities and the change in the objective function tend to zero. Therefore calculations with double precision are not sufficiently precise. Instead, we used the *mpf\_t* data type of the *GNU Multiple Precision Arithmetic Library*, which supplies arbitrary precision. Initially we start with a precision of 2000 bits as significant precision of the mantissa and increase the precision on demand, i.e., on every addition and subtraction we perform a check whether the current precision needs to be increased. The constants of PSO are assigned to values that are commonly used, i.e.,  $\chi = 0.72984$  and  $c_1 = c_2 = 1.496172$ , as proposed in [4]. The classical PSO is used as specified in Definition 1. Additionally, the PSO is visualized as pseudo code in Algorithm 1. No bound handling procedure is used. Particles move through the space without any borders. We checked our results with some benchmarks of [19] and an additional function. In detail, we investigate the sphere function, the high conditioned elliptic Function and Schwefel's problem (see [19] for detailed problem description). Additionally, the diagonal function is added, where its second derivation matrix has a single heavy eigenvalue and the corresponding eigenvector is oriented diagonally.

**Definition 12** (Objective functions). *The formal definitions of the used objective functions are*

- $f_{sph}(x) := \sum_{i=1}^D x_i^2$ , the sphere function,
- $f_{hce}(x) := \sum_{i=1}^D (10^6)^{\frac{i-1}{D-1}} x_i^2$ , the high conditioned elliptic function,
- $f_{sch}(x) := \sum_{i=1}^D (\sum_{j=1}^i x_j)^2$ , the Schwefel's problem and
- $f_{diag}(x) := \sum_{i=1}^D x_i^2 + 10^6 \cdot \left( \sum_{i=1}^D x_i \right)^2$ , the diagonal function.

The optimal position for all of these functions is the origin and the optimal value is zero.

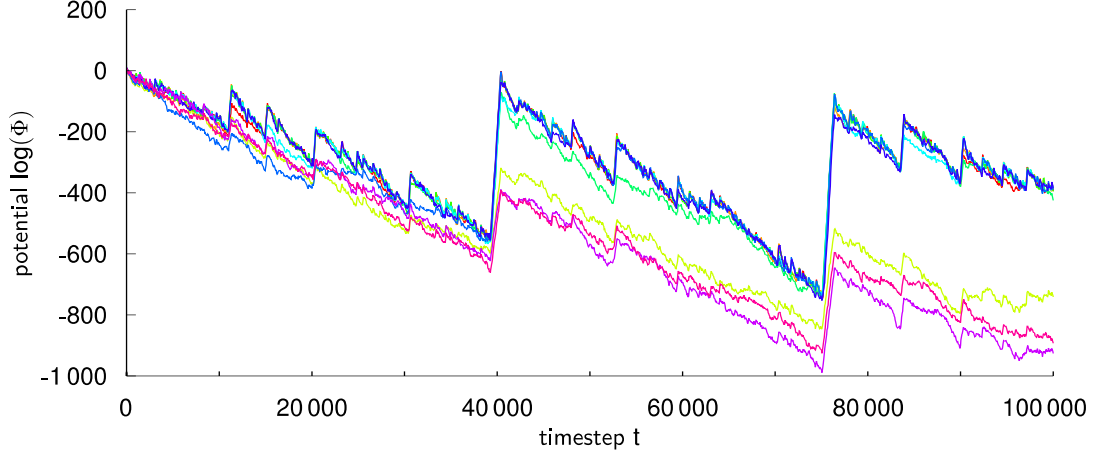
### 4.1 Used Software

For arbitrary precision the *mpf\_t* data type of the *GNU Multiple Precision Arithmetic Library* with version 4.3.2 is used. The program, which simulates PSO is implemented in C++ and uses the *mpf\_t* data type (program code is enclosed). The program runs on *openSUSE 12.3* and uses the *GNU Compiler Collection* with version 4.3.3. For creation of diagrams *plot2svg* of Jürg Schwizer within *Matlab R2013b* and *InkScape* 0.48 is used.

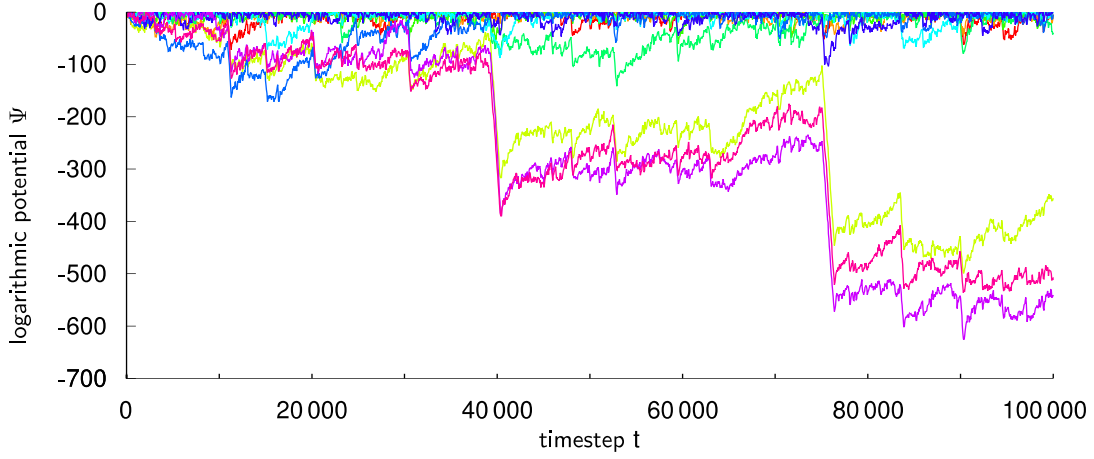
### 4.2 Expectation of increments for various scenarios

This section represents a motivation for stagnation phases with a specific number of stagnating dimensions. Firstly, some typical runs of the PSO are presented, which illustrate that for a combination of function and number of particles there is always a fixed number of dimensions, which are not stagnating. Secondly, the expectation of the increments are visualized to support this impression.

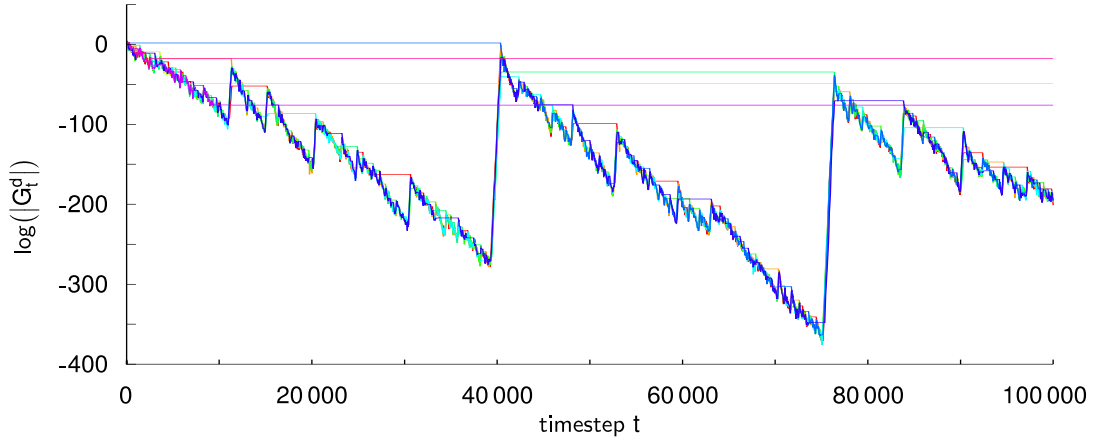
Figure 4 shows a typical run of the PSO on the sphere function with three particles and 10 dimensions. All positions are initialized uniformly at random in the 10-dimensional cube  $[-100.0, 100.0]^{10}$ . Additionally, zero velocity initialization is used, which means that all entries of all velocity vectors



(a) Logarithm of potential  $\log(\Phi)$

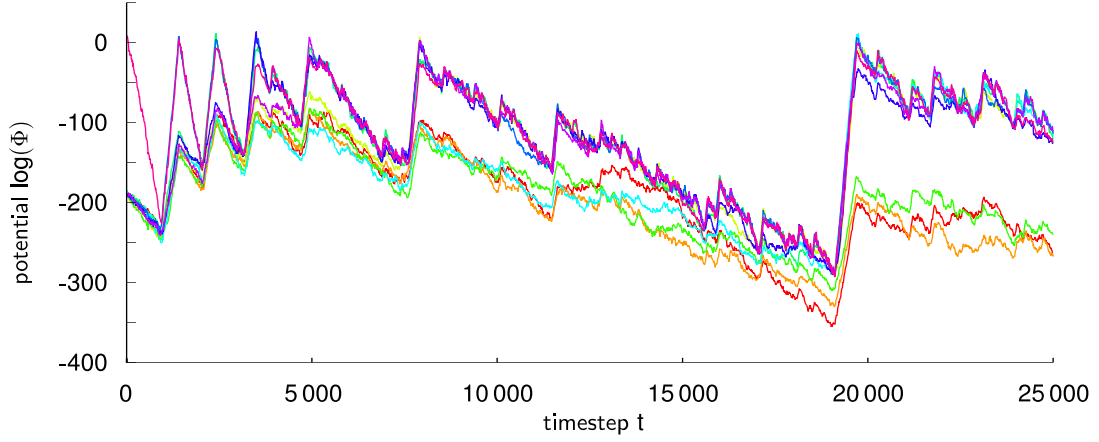


(b) Logarithmic potential  $\Psi$

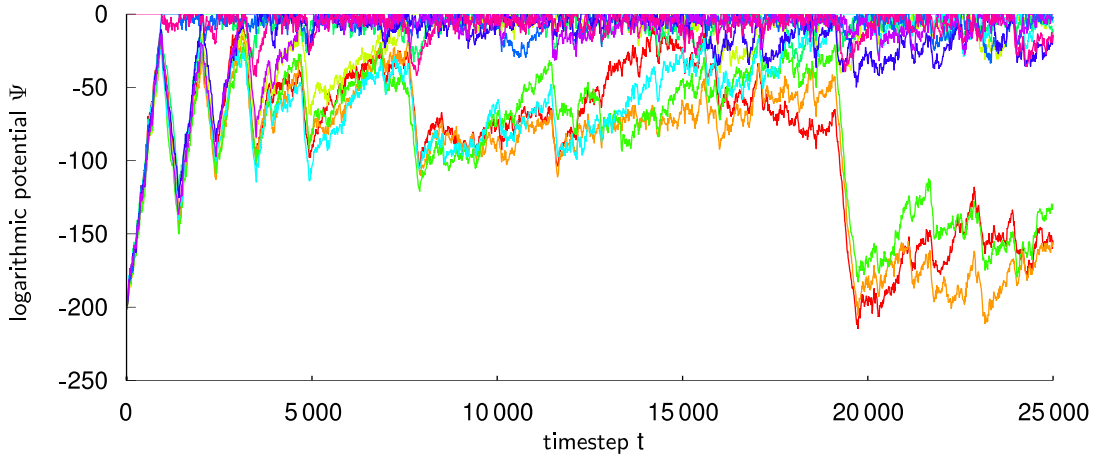


(c) Logarithm of distance to optimum

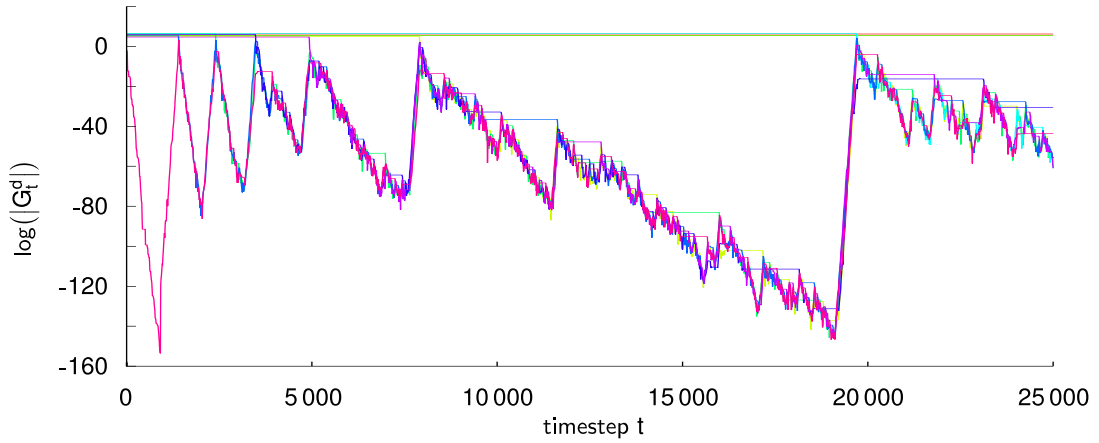
Figure 4: (a) Potential, (b) logarithmic potential and (c) distance to optimum with potential as specified in Definition 3 on sphere function in 10 dimensions and with 3 particles; each line represents a single dimension, usual initialization



(a) Logarithm of potential  $\log(\Phi)$



(b) Logarithmic potential  $\Psi$



(c) Logarithm of distance to optimum

Figure 5: (a) Potential, (b) logarithmic potential and (c) distance to optimum with potential as specified in Definition 3 on sphere function in 10 dimensions and with 3 particles; each line represents a single dimension, initialization with Algorithm 2

are initialized with the value zero. Other velocity initializations have also been investigated, but the results are the same. The initialization of positions uniformly at random in a cube and with zero velocity will be referred to as *usual initialization*, because this is a common initialization for PSO algorithms. The Figure 4 contains the logarithm of the potential, the logarithmic potential and the logarithm of the distance to the origin, which is the only optimal point. For some periods of time there are some dimensions which have low logarithmic potential and for the same dimensions the logarithm of the distance to the origin stays quite constant. If one of those dimensions regains potential then all dimensions regain potential, but dimensions which already have low logarithmic potential do increase much slower than the others. Therefore their logarithmic potential decreases. In Figure 4 approximately at time steps 40 000 and 75 000 a dimension, which had lower logarithmic potential than other dimensions, regains potential. At the same times the logarithmic potential of the remaining dimensions, which already had low logarithmic potential, significantly decreases. While some dimensions increase their potential in dimensions, which had large logarithmic potential previously, the respective positions may become worse. Finally, there are three dimensions which have a significantly smaller logarithmic potential than the other dimensions. The other 7 dimensions are optimizing. There are always short periods of time, where one of the 7 dimensions has a quite constant difference to the origin, as you can see in Figure 4. These periods can be that large that one of the three dimensions with low logarithmic potential regains potential, but as the logarithmic potential of the lower three dimensions continuously decreases, it becomes more and more unlikely that dimensions change from low logarithmic potential to high logarithmic potential. An interesting fact is that the number of optimizing dimensions in this experiments stay the same, even if the number of dimensions of the domain changes. For the sphere function and three particles finally  $\max(0, D - 7)$  dimensions appear, which have significantly smaller logarithmic potential than the other dimensions.

Figure 5 illustrates another run of the PSO on the sphere function with three particles and 10 dimensions. For this run the positions are initialized as specified in Algorithm 2. This algorithm first initializes all  $D$  dimensions as previously described uniformly at random in the interval  $[-100.0, 100.0]$ . The dimensions  $d^*$  to  $d^* + L - 1$  are initialized with some random center ( $Y \in [-100.0, 0, 100.0]$ ) and some random noise in the range  $[-100.0 \cdot 2^{-S}, 100.0 \cdot 2^{-S}]$ . This initialization simulates stagnation of  $L$  consecutive dimensions with initial logarithmic potential of approximately  $-S$ . The chosen parameters for the run visualized in Figure 5 are  $N = 3$  particles,  $D = 10$  dimensions, a scale of  $S = 200$ ,  $L = 9$  initial stagnating dimensions and  $d^* = 1$  the index of the first stagnating dimension.

Another option would have been scaling all position values of the first  $L$  dimensions with  $2^{-S}$ , but this is problematic. On the one hand the value of the initial potential then is mainly much less than  $-S$ , because not only the velocities are reduced, but also the derivative is smaller in the neighborhood of the optimum. The reduction of the derivative cannot be controlled with some general formula, because it depends on the objective function. On the other hand the logarithmic potential used in this paper on functions which are not composite, like  $f_{sch}$  and  $f_{diag}$ , acts not as proposed in the assumptions of Section 3 for this initialization. This is due to the fact that the derivative of those objective functions in stagnating dimensions depends also on the non-stagnating dimensions. If it changes which part overweights the other then also the distribution of the increments changes significantly during a stagnation phase. This change appears for the functions  $f_{sch}$  and  $f_{diag}$  when the positions of non-stagnating dimensions reach values in the interval  $[-100.0 \cdot 2^{-S}, 100.0 \cdot 2^{-S}]$ .

---

**Algorithm 2:** Special position initialization

---

**input** : number of particles  $N$ , number of dimensions  $D$ , scale  $S$ , number of dimensions with low logarithmic potential  $L$ , first dimension with low logarithmic potential  $d^*$   
**output**: a vector of initial positions for each particle  $X \in (\mathbb{R}^D)^N$

*/\* rand(a,b) supplies a uniform random value in [a,b] \*/*

```
1 for d := 1 to D do
2   for n := 1 to N do
3     | X[n][d] := rand(-100.0, 100.0);
4   end
5 end

6 for d := d* to d* + L - 1 do
7   Y := rand(-100.0, 100.0);
8   for n := 1 to N do
9     | X[n][d] := X[n][d] · 2-S + Y;
10  end
11 end

12 return X ;
```

---

**Example 1.** The function  $f(x, y) = x^2 + (x + y)^2$ , the Schwefel's problem in two dimensions, is analyzed. The first derivative equals  $(4x + 2y, 2x + 2y)$ . If  $y = -2$ , then the optimal value for  $x$  is 1, because then the first entry of the first derivative is zero. If  $x = 100 + 1$ , then all entries of the first derivate are mainly determined by the value of  $x$ . If the difference of  $x$  to the optimal value is decreased by a factor of 100 to  $x = 1 + 1$ , then the first derivative is also decreased approximately by a factor of 100. If the difference of  $x$  to the optimal value is again decreased by a factor of 100 to  $x = 0.01 + 1$ , then the first entry of the first derivative is also decrease by a factor of 100, but the second entry stays quite constant. Therefore the second entry of the first derivative is now mainly determined by  $y$ . The potential recognizes this change, too.

If the positions of the stagnating dimension would be initialized small compared to the other dimensions, then the position values of the stagnating dimensions stay constant while the position values of non-stagnating dimensions tend to their optimal value. Initially the complete first derivative is determined by the dimensions with larger position values. In the beginning the absolute values of all entries of the first derivative are decreasing, but finally the entries for stagnating dimensions of the first derivative stay constant. Therefore the behavior changes significantly if this border is crossed. For sure such situations may appear after initialization or during a PSO run, but it is very unlikely that the particles are initialized in that way if standard initializations are used and it is very unlikely that stagnating dimensions become much more optimized than non-stagnating dimensions during a PSO run. The objective functions  $f_{sph}$  and  $f_{hce}$  are not facing this problem, because the derivative of a single dimension depends only on the dimension itself.

For the PSO run in Figure 5 9 of 10 dimensions are initialized as stagnating dimensions with logarithmic potential  $-200$ . This figure illustrates that if less than seven dimensions are non-stagnating, then the logarithmic potential of stagnating dimensions is increasing. The smaller the number of non-stagnating dimensions is, the larger is the increase of the logarithmic potential.

Finally, one of the stagnating dimensions becomes significant. This results in a running phase in that dimension. Running phases are introduced in [16, 17, 18]. If the swarm is running in a dimension  $d$ , then the coordinate value of dimension  $d$  determines the local attractor and the influence of the other dimensions can be neglected. All velocities in direction  $d$  are either all positive or all negative, the local attractors are updated each step and the global attractor is updated at least once each iteration. Dimension  $d$  is heavily improved during this phase. The running phase terminates as soon as walking in dimension  $d$  does not lead to a further improvement of the function value. During this running phase, the previously non-stagnating dimensions and the running dimension regain potential. The other dimensions also experience an increase in potential, but this increase is much smaller compared to the running dimension and the non-stagnating dimensions. Therefore the logarithmic potential decreases. This behavior repeats until seven dimensions are non-stagnating. The largest running phase in Figure 5 occurs approximately at step 19 000.

The PSO runs visualized in Figures 4 and 5 do not look quite linear as proposed. This is due to the fact that they do not show continuous stagnation phases. Figure 6 represents the extended run of Figure 4, with hundred times more iterations. The linear drift of the logarithmic potential can easily be recognized in this figure. Furthermore, the non-stagnating dimensions can hardly be distinguished because their variance is not larger than in less number of iterations. As the scale has increased, the variance looks smaller.

**Definition 13.** *For an evaluated random variable of the  $r$ 'th test run, we write the original random variable in single tilted brackets with index  $r$ . For example  $\langle A \rangle_r$  is written for the evaluated random variable with name  $A$  in test run  $r$ .*

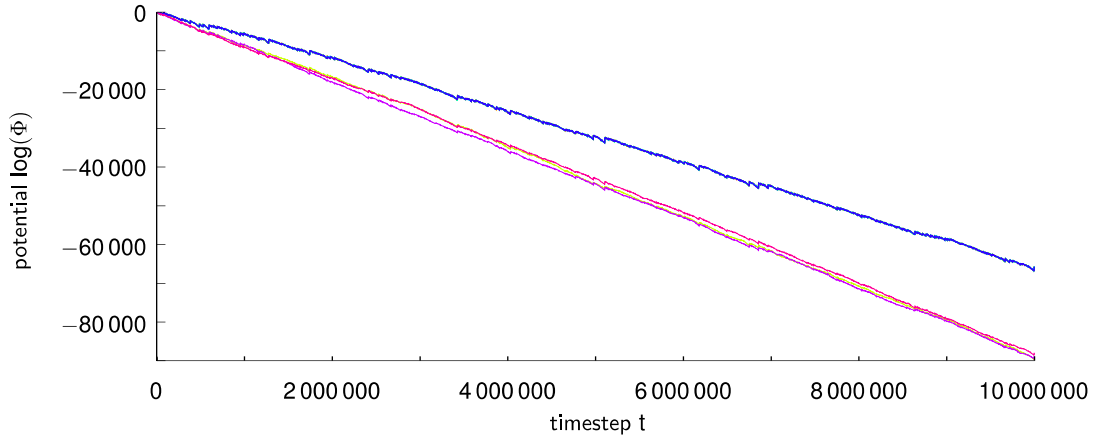
**Experiment 1.** *The first statistical analysis measures expected alteration of the logarithm of the potential and the logarithmic potential for various scenarios. All functions are evaluated with  $D = 10$  dimensions and 100 000 iterations. The stagnating dimensions, the number of particles  $N$  and the objective functions vary. Algorithm 2 is used for position initialization and all velocities are initially set to zero. The initial logarithmic potential is adjusted, such that for no test run the logarithmic potential of any stagnating dimension reaches a value of  $-100$  or more within the 100 000 iterations. Each tested configuration is started  $R = 500$  times with different seeds. Let  $D_S$  be the set of stagnating dimensions, which is fixed for each specific configuration. Let  $T_m$  be 50 000 and let  $T_e$  be 100 000, the number of half and full time steps of the evaluated process.*

The following estimators are defined:

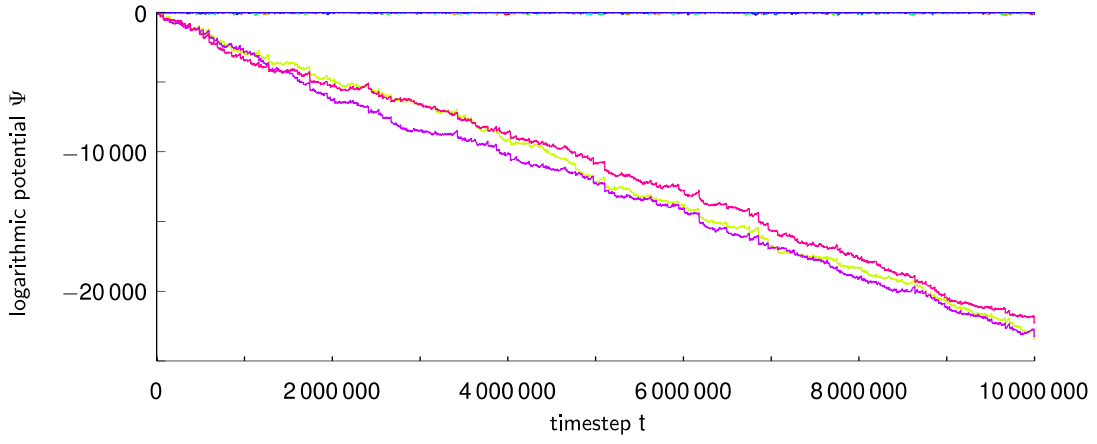
$$\begin{aligned}\overline{\mu_U} &:= \frac{1}{R(T_e - T_m)} \sum_{r=1}^R \left[ \log \left( \max_{d \notin D_S} \left( \left\langle \Phi \left( \frac{T_e}{\Delta t}, d \right) \right\rangle_r \right) \right) - \log \left( \max_{d \notin D_S} \left( \left\langle \Phi \left( \frac{T_m}{\Delta t}, d \right) \right\rangle_r \right) \right) \right], \\ \overline{\mu_M} &:= \frac{1}{R(T_e - T_m)} \sum_{r=1}^R \left[ \log \left( \min_{d \notin D_S} \left( \left\langle \Phi \left( \frac{T_e}{\Delta t}, d \right) \right\rangle_r \right) \right) - \log \left( \min_{d \notin D_S} \left( \left\langle \Phi \left( \frac{T_m}{\Delta t}, d \right) \right\rangle_r \right) \right) \right], \\ \overline{\mu_D} &:= \frac{1}{R \cdot |D_S| (T_e - T_m)} \sum_{r=1}^R \sum_{d \in D_S} \left[ \log \left( \left\langle \Phi \left( \frac{T_e}{\Delta t}, d \right) \right\rangle_r \right) - \log \left( \left\langle \Phi \left( \frac{T_m}{\Delta t}, d \right) \right\rangle_r \right) \right], \\ \overline{\mu_L} &:= \overline{\mu_D} - \overline{\mu_U}.\end{aligned}$$

As linearity is available only in logarithmic scale, the logarithm of the potential  $\Phi$  and the logarithmic potential  $\Psi$  is used.  $\overline{\mu_U}$  represents the average decrease or increase of the logarithm of

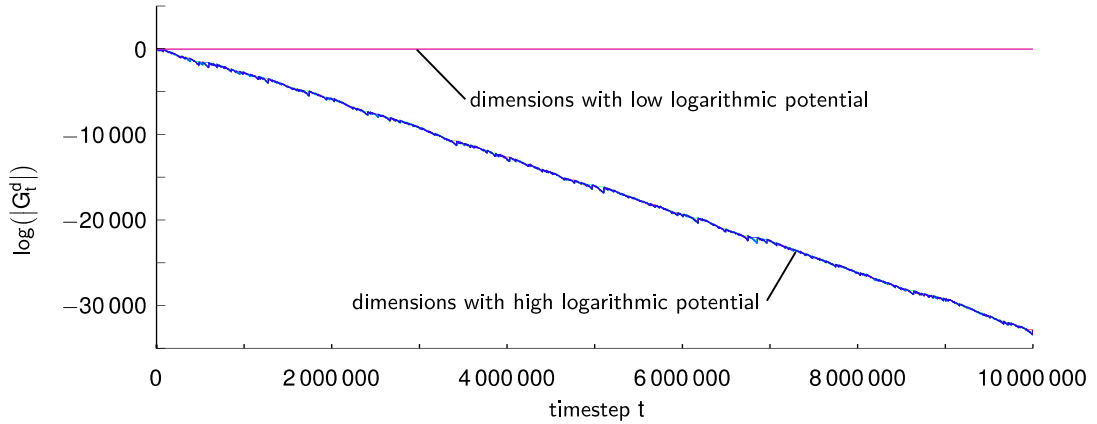




(a) Logarithm of potential  $\log(\Phi)$



(b) Logarithmic potential  $\Psi$



(c) Logarithm of distance to optimum

Figure 6: (a) Potential, (b) logarithmic potential and (c) distance to optimum with potential as specified in Definition 3 on sphere function in 10 dimensions and with 3 particles; each line represents a single dimension, usual initialization

the potential of the most significant dimension per time step,  $\overline{\mu_M}$  represents the average decrease or increase of the logarithm of the potential of the least significant dimension among the initially non-stagnating dimensions per time step,  $\overline{\mu_D}$  represents the average decrease or increase of the logarithm of the potential of stagnating dimensions per time step and  $\overline{\mu_L}$  represents the average decrease or increase of the logarithmic potential of the stagnating dimensions per time step. It is assumed that the first half of iterations is sufficient for adequate mixing. Then the second half is used for average calculation. Additionally the squared standard deviations are estimated for  $\overline{\mu_U}$ ,  $\overline{\mu_M}$ ,  $\overline{\mu_D}$  and  $\overline{\mu_L}$  by

$$\begin{aligned}\overline{\sigma_U^2} &:= \frac{1}{R(T_e - T_m)^2} \sum_{r=1}^R \left[ \log \left( \max_{d \notin D_S} \left( \left\langle \Phi \left( \frac{T_e}{\Delta t}, d \right) \right\rangle_r \right) \right) \right. \\ &\quad \left. - \log \left( \max_{d \notin D_S} \left( \left\langle \Phi \left( \frac{T_m}{\Delta t}, d \right) \right\rangle_r \right) \right) - (T_e - T_m) \overline{\mu_U} \right]^2, \\ \overline{\sigma_M^2} &:= \frac{1}{R(T_e - T_m)^2} \sum_{r=1}^R \left[ \log \left( \min_{d \notin D_S} \left( \left\langle \Phi \left( \frac{T_e}{\Delta t}, d \right) \right\rangle_r \right) \right) \right. \\ &\quad \left. - \log \left( \min_{d \notin D_S} \left( \left\langle \Phi \left( \frac{T_m}{\Delta t}, d \right) \right\rangle_r \right) \right) - (T_e - T_m) \overline{\mu_M} \right]^2, \\ \overline{\sigma_D^2} &:= \frac{1}{R \cdot |D_S| (T_e - T_m)^2} \sum_{r=1}^R \sum_{d \in D_S} \left[ \log \left( \left\langle \Phi \left( \frac{T_e}{\Delta t}, d \right) \right\rangle_r \right) \right. \\ &\quad \left. - \log \left( \left\langle \Phi \left( \frac{T_m}{\Delta t}, d \right) \right\rangle_r \right) - (T_e - T_m) \overline{\mu_D} \right]^2, \\ \overline{\sigma_L^2} &:= \frac{1}{R \cdot |D_S| (T_e - T_m)^2} \sum_{r=1}^R \sum_{d \in D_S} \left[ \left\langle \Psi \left( \frac{T_e}{\Delta t}, d \right) \right\rangle_r - \left\langle \Psi \left( \frac{T_m}{\Delta t}, d \right) \right\rangle_r - (T_e - T_m) \overline{\mu_L} \right]^2\end{aligned}$$

and  $\overline{\sigma_U} := \sqrt{\overline{\sigma_U^2}}$ ,  $\overline{\sigma_M} := \sqrt{\overline{\sigma_M^2}}$ ,  $\overline{\sigma_D} := \sqrt{\overline{\sigma_D^2}}$ ,  $\overline{\sigma_L} := \sqrt{\overline{\sigma_L^2}}$ . The variance will be analyzed in Section 4.4 in detail. In Figure 7 the measured decrease of the logarithm of the potential for non-stagnating dimension  $\overline{\mu_U}$ , for stagnating dimensions  $\overline{\mu_D}$ , the measured change in logarithmic potential  $\overline{\mu_L}$  and their measured standard deviations are visualized for the sphere function  $f_{sph}$ ,  $N = 3$  particles,  $D = 10$  dimensions and for different numbers of stagnating dimensions  $L$ . The decrease of the logarithm of the potential for non-stagnating dimensions  $\overline{\mu_U}$  becomes smaller as the number of non-stagnating dimensions increases. This is due to the fact that, the more dimensions need to be optimized, the more time is needed for optimization. The second columns represent the decrease of the logarithm of the potential for stagnating dimensions  $\overline{\mu_D}$ , which also becomes smaller, but not that fast as  $\overline{\mu_U}$ . The third columns, which represent the change of logarithmic potential  $\overline{\mu_L}$ , is heavily positive for only one non-stagnating dimension, but it decreases while the number of non-stagnating dimensions increases until finally the measured change becomes negative. If less than 7 dimensions are non-stagnating and further stagnating dimensions are available, then the logarithmic potential is likely to increase until at least one of the stagnating dimensions becomes non-stagnating, because  $\overline{\mu_L}$  is positive in those cases.

Figures 8 and 9 visualize the expectations of the change in logarithmic potential per time step for various scenarios. An obvious attribute of both graphs is that all functions perform similar

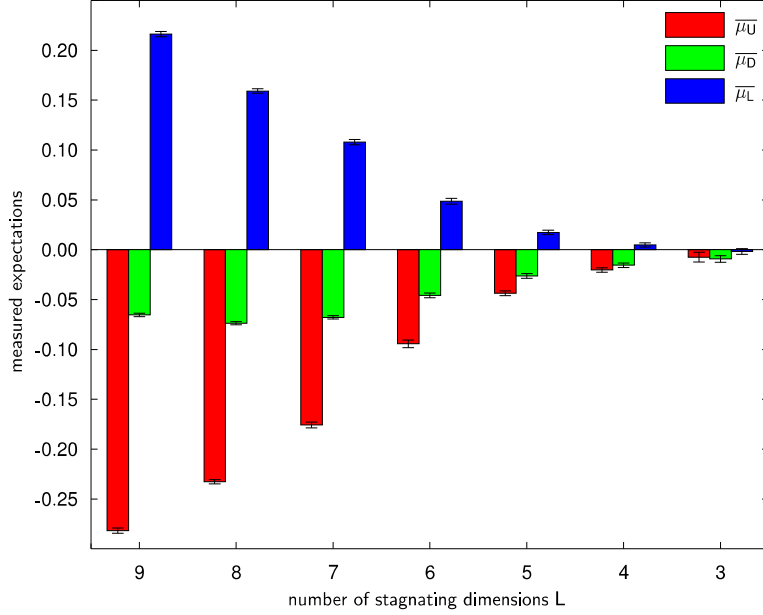


Figure 7: Measured expectations  $\overline{\mu_U}$ ,  $\overline{\mu_D}$  and  $\overline{\mu_L}$  and their standard deviations for sphere function  $f_{sph}$ ,  $N = 3$  particles,  $D = 10$  dimensions,  $d^* = 1$  and variable number of stagnating dimensions  $L$

if only one dimension is non-stagnating ( $L = 9$ ). All functions mentioned in this paper have a constant second derivative, which is a positive definite matrix, and third and higher derivatives are zero. If only a single dimension is not stagnating then the remaining function is represented by a parabola with some scale. Remaining function is meant as the function, which remains after all values of stagnating dimensions are replaced with the constants that represents their current position. The scale of the parabola does not influence the behavior of the PSO, because it only evaluates whether a position is better or not, which is not dependent on the scale. For sure the scale of a parabola has linear influence on  $\Phi$ , but the logarithmic potential is not affected, because this scale is compensated by the subtraction in Definition 2.

Furthermore, the sphere function  $f_{sph}$  and the high conditioned elliptic function  $f_{hce}$  have quite similar measured expectations. The high conditioned elliptic function actually is a sphere function scaled along the axes with different constant scales. It seems that the PSO does not care if such scaling is applied to the sphere function.

Other functions can have different behavior. For the Schwefel's problem it even depends on which dimensions are stagnating. To explain this property, examining the matrix representation of the mentioned objective functions is helpful:

- $f_{sph}(x) = x^t \cdot A_{sph} \cdot x$ , with  $(A_{sph})_{i,j} = 1$  if  $i = j$  and 0 else,
- $f_{hce}(x) = x^t \cdot A_{hce} \cdot x$ , with  $(A_{hce})_{i,j} = (10^6)^{\frac{i-1}{D-1}}$  if  $i = j$  and 0 else,
- $f_{sch}(x) = x^t \cdot A_{sch} \cdot x$ , with  $(A_{sch})_{i,j} = D + 1 - \min(i, j)$  and
- $f_{diag}(x) = x^t \cdot A_{diag} \cdot x$ , with  $(A_{diag})_{i,j} = 10^6 + (A_{sph})_{i,j}$ ,

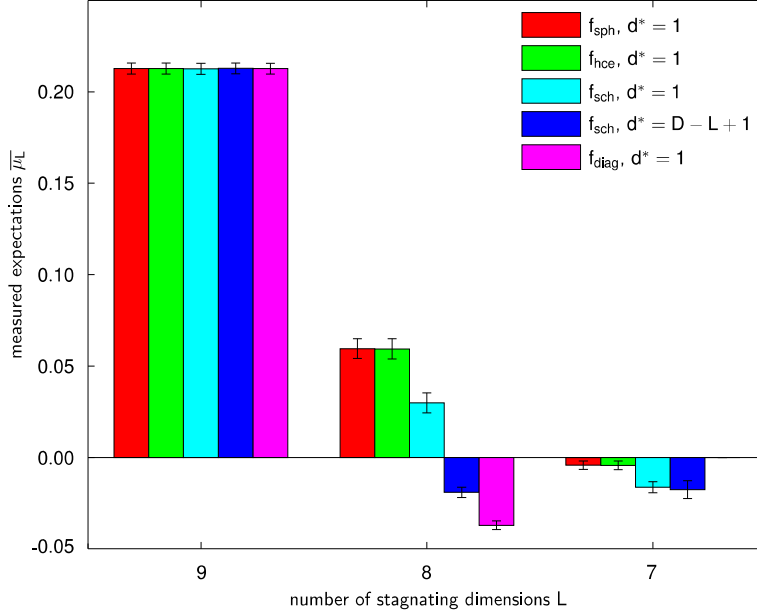


Figure 8: Measured expectations of logarithmic potential  $\overline{\mu_L}$  and their standard deviations for various functions,  $N = 2$  particles,  $D = 10$  dimensions and variable number of stagnating dimensions  $L$

where  $x^t$  represents the transposed vector of  $x$ . Important are the directions of the eigenvectors of those matrices. All eigenvectors have only real values, because the matrices are symmetric and positive semidefinite, which always is a property of second derivative matrices of objective functions, which are two times continuously differentiable, at local minima. For sphere function  $f_{sph}$  and high conditioned elliptic function  $f_{hce}$  there exists an eigenvector basis, which is parallel to the coordinate axes. This is the main reason for their similar behavior, because particles prefer to walk along the axes as proposed in [17]. In contrast, we have the diagonal function  $f_{diag}$  with matrix  $A_{diag}$ , which has a single large eigenvalue  $(D \cdot 10^6 + 1)$  and the corresponding eigenvector is oriented diagonally to all axes, i.e., the eigenvector is represented by  $(1, 1, \dots, 1)^t \in \mathbb{R}^D$ . The PSO has only bad performance on this function. For 2, 3 and 4 particles only two non-stagnating dimensions are necessary to generate considerable decreasing logarithmic potential, while for the sphere function and the high conditioned elliptic function there are 3 non-stagnating dimensions if 2 particles are available and there are 7 non-stagnating dimensions if 3 particles are available until the measured expectation of logarithmic potential becomes negative. For 4 or more particles operating on the sphere function or the high conditioned elliptic function it is even not known whether there exists a finite bound for the number of non-stagnating dimensions, such that the expectation of logarithmic potential becomes negative. As the logarithmic potential of the diagonal function  $f_{diag}$  is already negative for two non-stagnating dimensions, only the values for  $L = 9$  and  $L = 8$  are visualized in Figures 8 and 9. Negative expectation of logarithmic potential means that stagnating dimension prefer to decrease their logarithmic potential and therefore are more likely to stay stagnating.

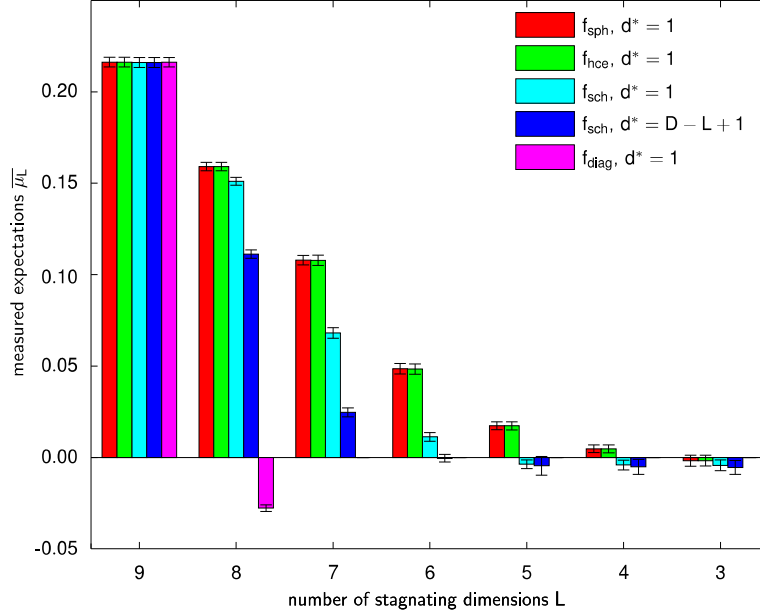


Figure 9: Measured expectations of logarithmic potential  $\overline{\mu_L}$  and their standard deviations for various functions,  $N = 3$  particles,  $D = 10$  dimensions and variable number of stagnating dimensions  $L$

If some dimensions are temporarily stagnating then the functions appear like

$$(x - x^*)^t \cdot A \cdot (x - x^*) + const,$$

where all rows and columns of stagnating dimensions are removed and  $x^*$  is the current optimal point for the remaining dimensions. For composite functions this optimal point stays at the origin. For other functions this is not necessarily the case. For the sphere function  $f_{sph}$  and the diagonal function  $f_{diag}$  the remaining submatrix does not depend on which dimensions are stagnating, but only on the number of stagnating dimensions. The submatrix of the high conditioned elliptic function  $f_{hce}$  depends on which dimensions are stagnating, but the remaining matrix is always a diagonal matrix and therefore the PSO always performs similar to the sphere function. This implies that for this objective function it also only depends on how many dimensions are currently stagnating.

For the Schwefel's problem it heavily depends on which dimensions are stagnating. E.g., two very different submatrices are the matrices where the first  $L$  dimensions are stagnating and where the last  $L$  dimensions are stagnating. The effect on the potential is visualized in the Figures 8, 9 and 10. If the first  $L$  dimensions are removed, then the remaining matrix is equal to the respective matrix of the Schwefel's problem in  $D - L$  dimensions. If the last  $L$  dimensions are removed, then the remaining matrix is equal to the respective matrix of the Schwefel's problem in  $D - L$  dimensions plus the value  $L$  for each entry for each entry of the remaining matrix. This is similar to the matrix of the diagonal function, where a value of  $10^6$  is added to each entry of the matrix of the sphere function. The more dimensions at the end are removed, the more intense becomes a

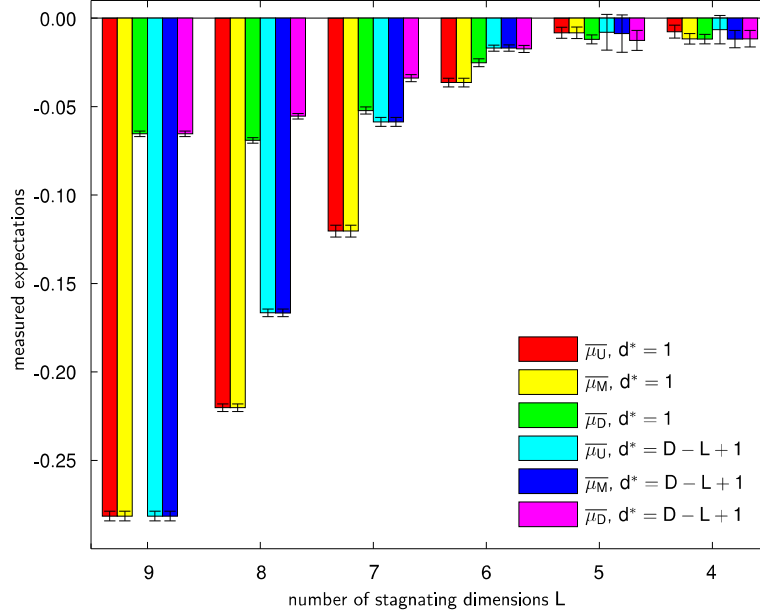


Figure 10: Measured expectations  $\overline{\mu_U}$ ,  $\overline{\mu_M}$  and  $\overline{\mu_D}$  and their standard deviations for Schwefel's problem  $f_{sch}$ ,  $N = 3$  particles,  $D = 10$  dimensions and variable number of stagnating dimensions  $L$

mostly diagonal eigenvector. Therefore the PSO performs much worse on the Schwefel's problem if the last  $L$  dimensions are stagnating than if the first  $L$  dimensions are stagnating. In Figures 8 and 9 this can be observed, because the change in logarithmic potential is much lower if the last  $L$  dimensions are stagnating.

Figure 10 even shows that the final number of stagnating dimensions can differ if different dimensions are stagnating. It shows the measured expectations of the change in potential of the largest potential value, the lowest potential value of the initial non-stagnating dimensions and the average of the initial stagnating dimensions. If the number of non-stagnating dimensions is small then  $\overline{\mu_M}$  is similar to  $\overline{\mu_U}$ , which means that the  $D - L$  initially non-stagnating dimensions stay non-stagnating. Finally the value of  $\overline{\mu_M}$  is similar to  $\overline{\mu_D}$ , which means that at least one of the initially non-stagnating dimensions becomes stagnating. Actually this change should happen as soon as  $\overline{\mu_U}$  is already larger than  $\overline{\mu_D}$  for fewer non-stagnating dimensions, because then there exist a dimension such that if this dimension becomes also stagnating the logarithmic potential of stagnating dimensions is more likely to decrease. In Figure 10  $\overline{\mu_U}$  is larger than  $\overline{\mu_D}$  for  $d^* = 1$  and  $L = 5$ . As expected  $\overline{\mu_M}$  is similar to  $\overline{\mu_D}$  for  $d^* = 1$  and  $L < 5$ . For  $d^* = D - L + 1$  this is not that clear.  $\overline{\mu_U}$  is slightly larger than  $\overline{\mu_D}$  for  $L = 6$ , but as the resulting decrease of logarithmic potential is very small, it may happen through random fluctuation that all initially non-stagnating dimensions regain potential from time to time. Therefore  $\overline{\mu_M}$  only slightly differs from  $\overline{\mu_U}$  for  $L = 5$ . If the number of iterations is heavily increased, then it will be more likely that one of the initial non-stagnating dimensions becomes stagnating.

Detailed results on the exact values of  $\overline{\mu_U}$ ,  $\overline{\mu_M}$ ,  $\overline{\mu_D}$  and  $\overline{\mu_L}$  of Experiment 1 are listed in Appendix A.

### 4.3 Verification of the Assumption of Separation

In this section evidence is provided for the separation of the increments  $I_{t,d}$  into the two parts  $B_{t,\Gamma_{B,T_0}}$  and  $J_{t,d,\Gamma_{J,T_0}}$ , as proposed in Assumption 1. The following experiment supplies information about the PSO during stagnation phases.

**Experiment 2.** *All functions are evaluated with  $D = 100$  dimensions and 200 000 iterations. The number of particles  $N$  and the objective functions vary. Algorithm 2 is used for position initialization and all velocities are initially set to zero.  $L$ , the number of initially stagnating dimensions, and  $d^*$ , the index of the first stagnating dimension, are chosen, such that the expectation of the increments of stagnating dimensions are just becoming negative.  $S$ , the initial scale of the stagnating dimensions, is set to 500. For no test run the logarithmic potential of any stagnating dimension reaches a value of  $-100$  or more within the 200 000 iterations. Each tested configuration is started  $R = 500$  times with different seeds. Let  $D_S$  be the set of stagnating dimensions, which is fixed for each specific configuration. Let  $T_m$  be 100 000 and let  $T_e$  be 200 000, the number of half and full time steps of the evaluated process.  $\Delta t$  is set to one for this experiment.*

The number of dimensions is taken that large, such that it can be assumed that  $B_{t,\Gamma_{B,T_0}}$ , the base increment of all stagnating dimensions, can be approximated by the average of all increments of stagnating dimensions. The initialization with Algorithm 2 and specific parameters  $L$  and  $d^*$  leads to a fixed set of stagnating dimensions. As discussed earlier, the set of stagnating dimensions fixes the expectation of the increments, but the set of stagnating dimensions also fixes the distributions of the increments, because the remaining function is always the same and therefore the PSO performs similar. Therefore it is assumed that  $\Gamma_{B,T_0}$  and  $\Gamma_{J,T_0}$  are constant distributions for a single scenario. For reasons of readability, the constant distributions are removed from the index of the estimators for  $B$  and  $J$ . The first half of the iterations is again used for sufficient mixing. With this experiment large sums of increments are investigated. Therefore  $\Delta t$  is set to one for this experiment, because larger values of  $\Delta t$  are implicitly shown with larger numbers of accumulated increments. Furthermore no scaling of the time axis is existent with this value of  $\Delta t$ . The following estimators for  $B$  and  $J$  in the  $r$ 'th run and time step  $t$  are used:

$$\begin{aligned}\overline{\langle B_t \rangle_r} &:= \sum_{d \in D_S} \langle I_{t,d} \rangle_r / |D_S| \\ \overline{\langle J_{t,d} \rangle_r} &:= \langle I_{t,d} \rangle_r - \overline{\langle B_t \rangle_r}\end{aligned}$$

Additionally the following helping variables for sums of  $\tau$  values are defined:

$$\begin{aligned}\langle I'_{\tau,d} \rangle_r &:= \sum_{t=T_m}^{T_m+\tau-1} \langle I_{t,d} \rangle_r \\ \overline{\langle B'_\tau \rangle_r} &:= \sum_{t=T_m}^{T_m+\tau-1} \overline{\langle B_t \rangle_r} \\ \overline{\langle J'_{\tau,d} \rangle_r} &:= \sum_{t=T_m}^{T_m+\tau-1} \overline{\langle J_{t,d} \rangle_r}\end{aligned}$$

The estimators for the expectation are calculated over the complete second half to receive a reliable value as in the previous experiment:

$$\overline{\mu_I} := \overline{\mu_B} := \frac{1}{R \cdot (T_e - T_m)} \sum_{r=1}^R \sum_{t=T_m}^{T_e-1} \overline{\langle B_t \rangle_r}$$

As proposed in Assumption 1 the measured expectation of  $J$  is zero:

$$\begin{aligned} \overline{\mu_J} &:= \frac{1}{R \cdot |D_S| \cdot (T_e - T_m)} \sum_{r=1}^R \sum_{d \in D_S} \sum_{t=T_m}^{T_e-1} \overline{\langle J_{t,d} \rangle_r} \\ &= \frac{1}{R \cdot |D_S| \cdot (T_e - T_m)} \cdot \sum_{r=1}^R \sum_{t=T_m}^{T_e-1} \sum_{d \in D_S} \left( \langle I_{t,d} \rangle_r - \sum_{\tilde{d} \in D_S} \langle I_{t,\tilde{d}} \rangle_r / |D_S| \right) = 0 \end{aligned}$$

A further estimator is introduced to measure the covariance ( $cov$ ) of the sums  $\sum_{t=T_m}^{T_m+\tau-1} B_t$  and  $\sum_{t=T_m}^{T_m+\tau-1} J_{t,d}$ :

$$\overline{cov_{B,J,\tau}} := \frac{1}{R \cdot |D_S|} \sum_{r=1}^R \sum_{d \in D_S} \left( \overline{\langle B'_\tau \rangle_r} - \tau \cdot \overline{\mu_B} \right) \cdot \overline{\langle J'_{\tau,d} \rangle_r}.$$

If  $B$  and  $J$  are independent then their covariance needs to be zero. Actually the estimator  $\overline{cov_{B,J,\tau}}$  for this covariance always evaluates to zero, because

$$\begin{aligned} \overline{cov_{B,J,\tau}} &:= \frac{1}{R \cdot |D_S|} \sum_{r=1}^R \sum_{d \in D_S} \left[ \left( \overline{\langle B'_\tau \rangle_r} - \tau \cdot \overline{\mu_B} \right) \cdot \overline{\langle J'_{\tau,d} \rangle_r} \right] \\ &= \frac{1}{R \cdot |D_S|} \sum_{r=1}^R \left[ \left( \overline{\langle B'_\tau \rangle_r} - \tau \cdot \overline{\mu_B} \right) \cdot \sum_{d \in D_S} \left( \langle I'_{\tau,d} \rangle_r - \overline{\langle B'_\tau \rangle_r} \right) \right] \\ &= \frac{1}{R \cdot |D_S|} \sum_{r=1}^R \left[ \left( \overline{\langle B'_\tau \rangle_r} - \tau \cdot \overline{\mu_B} \right) \cdot \left( \left( \sum_{d \in D_S} \langle I'_{\tau,d} \rangle_r \right) - \underbrace{|D_S| \cdot \overline{\langle B'_\tau \rangle_r}}_{=\sum_{d \in D_S} \langle I'_{\tau,d} \rangle_r} \right) \right] = 0. \end{aligned}$$

Furthermore, this idea can be expanded. We use

$$\mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d \in D_S} I_{t,d} \mid \mathcal{A}'_t \right] \text{ with } \mathcal{A}'_t := \sigma \left( \mathcal{A}_t, \sigma \left( ((s_t^n)^{d \notin D_S}, (r_t^n)^{d \notin D_S})_{n \in \{1, \dots, N\}} \right) \right)$$

as approximation on the base increment  $B_t$ . This conditional expectation represents the average increment of stagnating dimensions, after the movement in non-stagnating dimensions is determined, i. e.,  $(s_t^n)^{d \notin D_S}$  and  $(r_t^n)^{d \notin D_S}$ , the random variables of the movement equations defined in Definition 1 for non-stagnating dimensions, and all information about previous steps are known. Even for the sum of  $\tau$  increments

$$\mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d} \mid \mathcal{A}_{T_m}^\tau \right] \text{ with } \mathcal{A}_{T_m}^\tau := \sigma \left( \mathcal{A}_{T_m}, ((s_t^n)^{d \notin D_S}, (r_t^n)^{d \notin D_S})_{\substack{n \in \{1, \dots, N\}, \\ t \in \{T_m, \dots, T_m+\tau-1\}}} \right)$$



is a possible approximation on  $\sum_{t=T_m}^{T_m+\tau-1} B_t$  and

$$\sum_{t=T_m}^{T_m+\tau-1} I_{t,d} - \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d} \middle| \mathcal{A}_{T_m}^\tau \right]$$

is a possible approximation on  $\sum_{t=T_m}^{T_m+\tau-1} J_{t,d}$ . Other choices for  $\mathcal{A}_t'$  and  $\mathcal{A}_{T_m}^\tau$  are also possible, but the inclusions  $\mathcal{A}_t \subseteq \mathcal{A}_t' \subseteq \mathcal{A}_{t+1}$  and  $\mathcal{A}_{T_m} \subseteq \mathcal{A}_{T_m}^\tau \subseteq \mathcal{A}_{T_m+\tau}$  need to be satisfied to grant the specified measurability of Assumption 1. With similar calculations as with the estimators, the actual expectation of  $J_t$  and the covariance of random variables  $B_t$  and  $J_t$  is zero, if the proposed approximations on random variables  $B$  are accepted. As presented in Example 1 the derivative in stagnating dimensions becomes quite constant if the coordinate values of stagnating dimensions become constant. For Figures 4, 5 and 6 it is significant that even temporarily stagnating dimensions do not change their position values significantly while they are stagnating. As the logarithm of the potential of all dimensions tends to minus infinity, as presented in Figure 6a, the absolute values of the velocities in all dimensions need to tend to zero. Therefore the potential in stagnating dimensions can be calculated by the multiplication of the absolute value of the velocity in that dimension and the approximately constant derivative in that dimension. As the increments are calculated by the difference of logarithmic potential values, only the portion of the velocity and the portion of the dimension with maximal potential remains. Hence the increments of all stagnating dimensions are only determined by the velocities in that dimension. The portion of the dimension with maximal potential is completely included in the base increment and therefore the dimension dependent part  $J$  of the increments is only determined by the change of the velocity in the specific dimension. The update of global and local attractors appear simultaneously for each dimension. Hence all dimension dependent parts  $J$  have quite the same distribution and are quite independent of each other, because the velocity change is determined only by the updates of the attractors and the two random variables of the movement equations, which belong to that specific dimension. That is the reason for averaging the expectation of random variables of type  $J$  and the covariance of random variables  $B$  and  $J$  over all stagnating dimensions. Therefore  $\mu_J$ , the actual expectation for a random variable  $J$ , is calculated with the previously described approximation by the following equations:

$$\mathbb{E} \left[ \sum_{t=T_m}^{T_m+\tau-1} J_{t,d} \middle| \mathcal{A}_{T_m} \right]$$

The variable  $J_{t,d}$  is replaced by  $I_{t,d} - B_t$  and then the approximation on  $B_t$  is applied.

$$\approx \frac{1}{|D_S|} \sum_{d \in D_S} \mathbb{E} \left[ \sum_{t=T_m}^{T_m+\tau-1} I_{t,d} - \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \middle| \mathcal{A}_{T_m}^\tau \right] \middle| \mathcal{A}_{T_m} \right]$$

Additivity of the expected value is applied. In the second part then only  $\sigma$ -algebra  $\mathcal{A}_{T_m}$  remains, because it is a sub- $\sigma$ -algebra of  $\mathcal{A}_{T_m}^\tau$ , i. e.,  $\mathcal{A}_{T_m} \subseteq \mathcal{A}_{T_m}^\tau$ .

$$\mathcal{A}_{T_m} \subseteq \mathcal{A}_{T_m}^\tau \implies \frac{1}{|D_S|} \sum_{d \in D_S} \mathbb{E} \left[ \sum_{t=T_m}^{T_m+\tau-1} I_{t,d} \middle| \mathcal{A}_{T_m} \right] - \frac{|D_S|}{|D_S|} \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \middle| \mathcal{A}_{T_m} \right] = 0.$$

Additionally  $cov_{B,J,\tau}$ , the actual covariance of random variables  $B$  and  $J$ , is calculated by

$$cov_{B,J,\tau} := \frac{1}{|D_S|} \sum_{d \in D_S} \mathbb{E} \left[ \left( \sum_{t=T_m}^{T_m+\tau-1} B_t - \mathbb{E} \left[ \sum_{t=T_m}^{T_m+\tau-1} B_t \mid \mathcal{A}_{T_m} \right] \right) \cdot \left( \sum_{t=T_m}^{T_m+\tau-1} J_{t,d} \mid \mathcal{A}_{T_m} \right) \right] \approx$$

All variables  $J_{t,d}$  and  $B_t$  are replaced by their approximation.

$$\approx \frac{1}{|D_S|} \sum_{d \in D_S} \mathbb{E} \left[ \left( \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \mid \mathcal{A}_{T_m}^\tau \right] - \mathbb{E} \left[ \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \mid \mathcal{A}_{T_m}^\tau \right] \mid \mathcal{A}_{T_m} \right] \right) \cdot \left( \sum_{t=T_m}^{T_m+\tau-1} I_{t,d} - \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \mid \mathcal{A}_{T_m}^\tau \right] \right) \mid \mathcal{A}_{T_m} \right]$$

The two nested conditional expectations in the second line can be combined and only  $\mathcal{A}_{T_m}$  remains, because  $\mathcal{A}_{T_m}^\tau$  is a sub- $\sigma$ -algebra of  $\mathcal{A}_{T_m}$ . Furthermore the outmost conditional expectation is replaced by two conditional expectations, where the inner  $\sigma$ -algebra  $\mathcal{A}_{T_m}^\tau$  is a sub- $\sigma$ -algebra of  $\mathcal{A}_{T_m}$ .

$$= \frac{1}{|D_S|} \sum_{d \in D_S} \mathbb{E} \left[ \mathbb{E} \left[ \left( \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \mid \mathcal{A}_{T_m}^\tau \right] - \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \mid \mathcal{A}_{T_m} \right] \right) \cdot \left( \sum_{t=T_m}^{T_m+\tau-1} I_{t,d} - \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \mid \mathcal{A}_{T_m}^\tau \right] \right) \mid \mathcal{A}_{T_m}^\tau \right] \mid \mathcal{A}_{T_m} \right]$$

The second outmost conditional expectation is applied. The first factor is a constant under this  $\sigma$ -algebra. Therefore nothing happens there. In the second factor the conditional expectation only remains for the sum over  $I_{t,d}$ .

$$= \frac{1}{|D_S|} \sum_{d \in D_S} \mathbb{E} \left[ \left( \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \mid \mathcal{A}_{T_m}^\tau \right] - \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \mid \mathcal{A}_{T_m} \right] \right) \cdot \left( \mathbb{E} \left[ \sum_{t=T_m}^{T_m+\tau-1} I_{t,d} \mid \mathcal{A}_{T_m}^\tau \right] - \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \mid \mathcal{A}_{T_m}^\tau \right] \right) \mid \mathcal{A}_{T_m} \right]$$

The sum  $\frac{1}{|D_S|} \sum_{d \in D_S}$  in front of the complete expression is shifted into the conditional expectation. The first factor does not depend on  $d$ . Therefore it can be applied directly on the second factor. On the one hand the sum is put inside the first conditional expectation of the second factor. On the other hand the second conditional expectation does not depend on  $d$  and therefore it is summed up  $|D_S|$  times, which means that this conditional expectation is effectively multiplied by one.

$$= \mathbb{E} \left[ \left( \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \mid \mathcal{A}_{T_m}^\tau \right] - \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \mid \mathcal{A}_{T_m} \right] \right) \cdot \underbrace{\left( \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d} \mid \mathcal{A}_{T_m}^\tau \right] - \mathbb{E} \left[ \frac{1}{|D_S|} \sum_{d' \in D_S} \sum_{t=T_m}^{T_m+\tau-1} I_{t,d'} \mid \mathcal{A}_{T_m}^\tau \right] \right)}_{=0} \mid \mathcal{A}_{T_m} \right] = 0.$$

Admittedly, this calculations can also be done for any other set of random variables, which does not belong to a stagnation phase at all, but then the described simplification cannot be justified. As all stagnating dimensions act similar and can be permuted without any effect, also the covariances of  $J_{t,d}$  for different values of  $d$  are zero. Therefore we have that there definitely exist random variables  $B$  and  $J$ , which have no linear dependency. Furthermore with the chosen approximation for variables base increment and dimension dependent increment we can expect sufficient independence for specific times or time intervals.

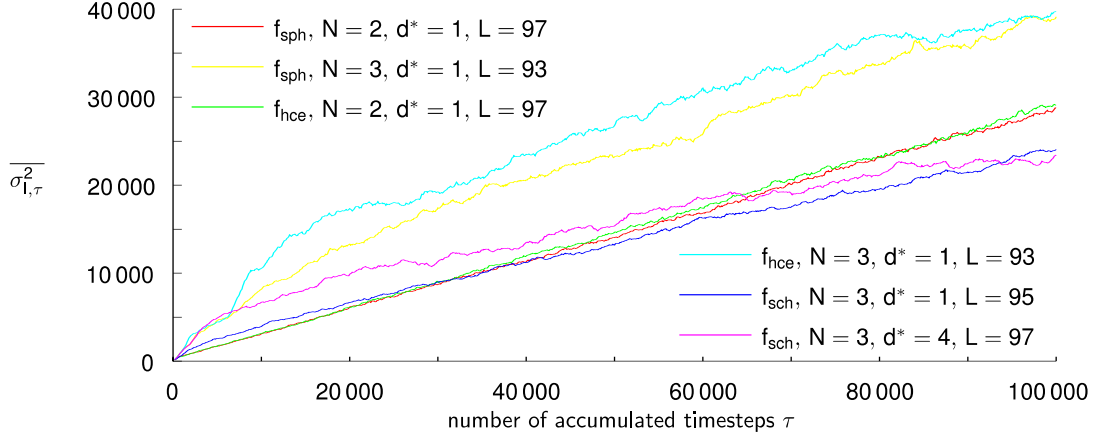
#### 4.4 Experimental Verification of the Assumption of Independence

The previous section presented reasons for the separation of the increments into independent base increments and dimension dependent increments. This section experimentally shows that increments, which belong to different periods of time, fulfill sufficient independence to accept the applicability of the theorems in Chapter 3.

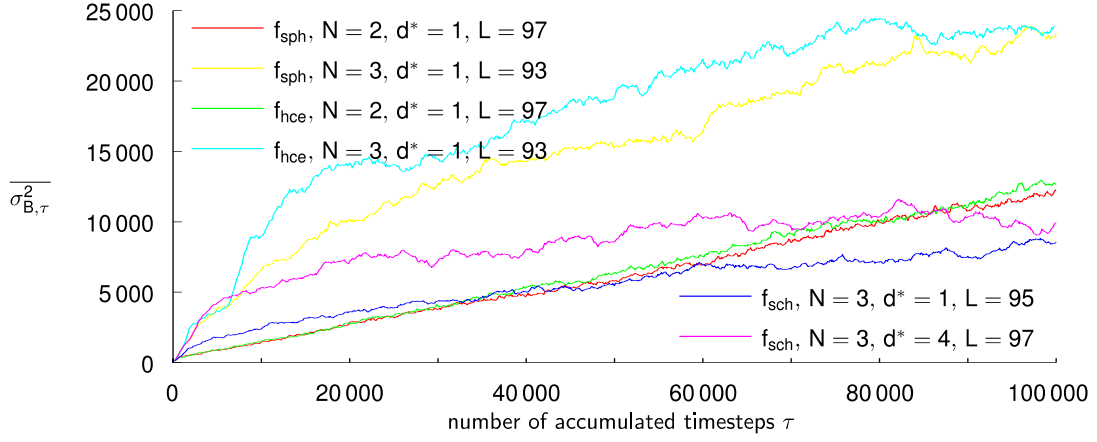
First some additional estimaters for the variance and the sixth central moment related to Experiment 2 are introduced.

$$\begin{aligned}\overline{\sigma_{I,\tau}^2} &:= \frac{1}{R \cdot |D_S|} \sum_{r=1}^R \sum_{d \in D_S} (\langle I'_{\tau,d} \rangle_r - \tau \cdot \overline{\mu_I})^2 \\ \overline{\sigma_{B,\tau}^2} &:= \frac{1}{R} \sum_{r=1}^R (\langle B'_\tau \rangle_r - \tau \cdot \overline{\mu_B})^2 \\ \overline{\sigma_{J,\tau}^2} &:= \frac{1}{R \cdot |D_S|} \sum_{r=1}^R \sum_{d \in D_S} \langle J'_{\tau,d} \rangle_r^2 \\ \overline{M_{I,\tau}^6} &:= \frac{1}{R \cdot |D_S|} \sum_{r=1}^R \sum_{d \in D_S} (\langle I'_{\tau,d} \rangle_r - \tau \cdot \overline{\mu_I})^6 \\ \overline{M_{B,\tau}^6} &:= \frac{1}{R} \sum_{r=1}^R (\langle B'_\tau \rangle_r - \tau \cdot \overline{\mu_B})^6 \\ \overline{M_{J,\tau}^6} &:= \frac{1}{R \cdot |D_S|} \sum_{r=1}^R \sum_{d \in D_S} \langle J'_{\tau,d} \rangle_r^6\end{aligned}$$

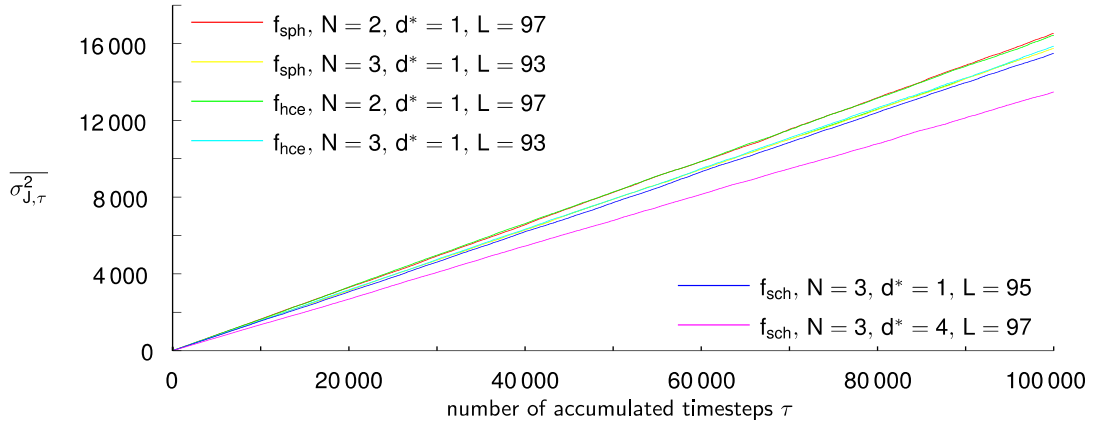
Due to limited resources only the sphere function  $f_{sph}$  with  $N = 2$  and  $N = 3$  particles, the high conditioned elliptic function  $f_{hce}$  with  $N = 2$  and  $N = 3$  particles and the Schwefel's problem  $f_{sch}$  with  $N = 3$  particles and two different sets of stagnating dimensions are investigated. As presented in Figures 8 and 9, the first negative value for the expectation of the increments is available with  $L = D - 3 = 97$  stagnating dimensions with two particles and sphere or high conditioned elliptic function,  $L = D - 7 = 93$  stagnating dimensions with three particles and sphere or high conditioned elliptic function and  $L = D - 5 = 95$  stagnating dimensions and  $d^* = 1$  with three particles and Schwefel's problem. If  $d^*$  is not equal to one then the remaining function of the Schwefel's problem depends heavily on the number of dimensions. For  $D = 10$  dimensions the suitable values are  $d^* = 5$  and  $L = 6$ , but for  $D = 100$  dimensions already  $d^* = 4$  and  $L = 97$  are the appropriate values. The measured variances are visualized in Figure 11. If the increments would



(a) Estimator for the variance of increments



(b) Estimator for the variance of base increments



(c) Estimator for the variance of dimension dependent increments

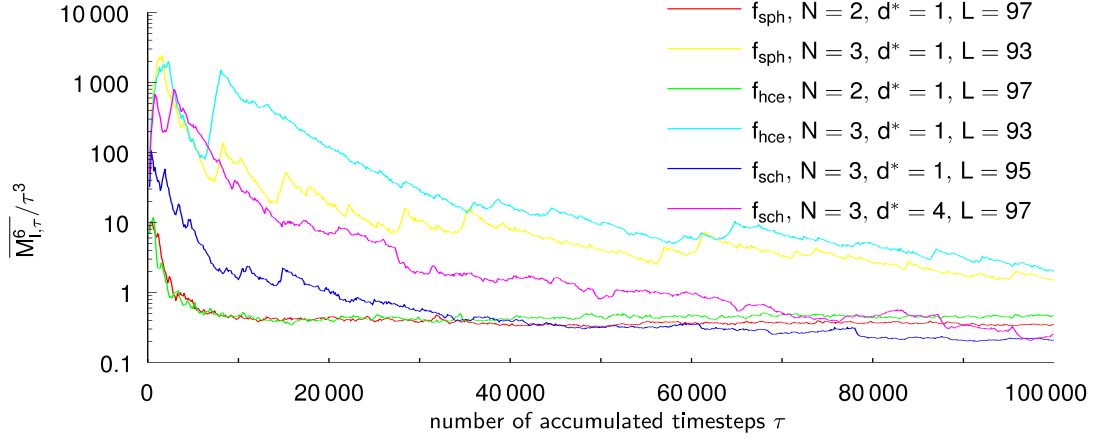
Figure 11: Estimators for the variance of (a) increments  $\overline{\sigma_{I,\tau}^2}$ , (b) base increments  $\overline{\sigma_{B,\tau}^2}$  and (c) dimension dependent increments  $\overline{\sigma_{J,\tau}^2}$  resulting from Experiment 2; each line represents a single configuration

be completely independent then the resulting graphs would be just lines. As already discussed, the distribution of the increments depends heavily on the number of non-stagnating dimensions. In Figure 4c at approximately steps 40 000 to 75 000 there are temporary four dimensions, which do not improve their position, but for the sphere function and three particles there generally are seven non-stagnating dimensions. Therefore with  $D = 10$  dimensions only three stagnating dimensions normally appear, because otherwise the expectation of the increments of stagnating dimensions is then positive. The additional temporary stagnating dimension has a slightly lower logarithmic potential than the other non-stagnating dimensions, which can be recognized in Figure 4b. Hence the logarithmic potential of the three obviously stagnating dimensions have a positive drift during this period of time with a temporary additional stagnating dimension. The temporary stagnating dimension also receives that positive drift, but as this is a random process, the logarithmic potential of this dimension can remain small for some time. At the end of periods of time with temporary additional stagnating dimensions the PSO becomes running in the temporary stagnating dimension and the other stagnating dimensions quickly lose logarithmic potential. Such periods appear quite frequently during PSO runs and their lengths vary. The fewer the difference between the expectation of the increments with current number of stagnating dimensions and with one dimension fewer is, the longer those periods can become. The heavier increase of the measured variances in Figure 11 results from those periods, because during or at the end of those periods the behavior differs. Over larger periods this effect become less important. Also Figure 11 supports that impression. After the heavy increase at the beginning, the variances grow quite linearly. For the sphere and high conditioned elliptic function with two particles, there is not even a recognizable increase at the beginning, because the difference of the expectation of the increments differ strongly for  $L = D - 3$  and  $L = D - 2$  dimensions (compare with Figure 8) and therefore the lengths of periods with temporary additional stagnating dimensions are very small. Altogether increments may be independent enough in time when large time scales are used.

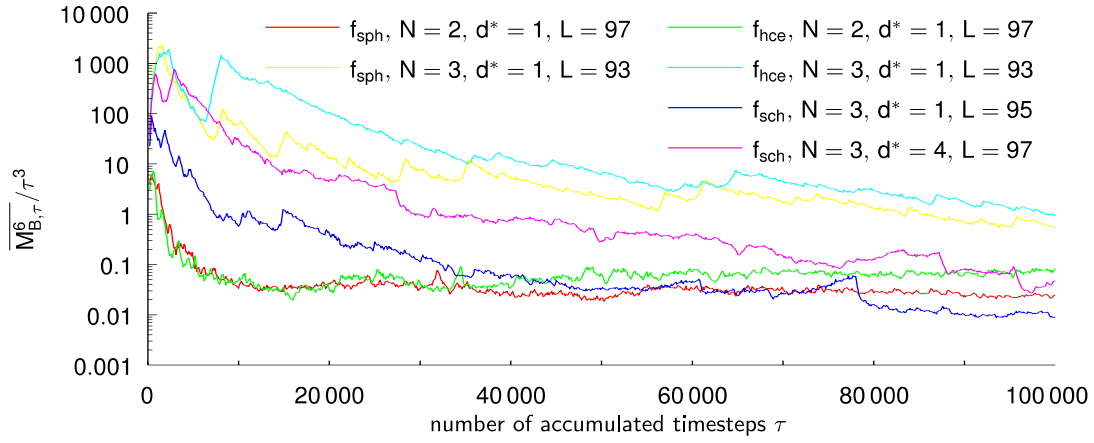
Taking a deeper look into the proof of Lemma 2 results in the awareness that the independence in  $t$  is used three times. It is used to guarantee that,

- $E[(\sum_{i=0}^{t-1}(I_i - \mu^*))^6] \leq C \cdot t^3$ , where  $C$  is some constant and  $I$  is either the base increment plus drift or the dimension dependent increment plus drift,
- there is a positive probability for finite times that the sum  $\sum_{i=0}^{t-1} I_i$  stays negative and
- the probability to receive a negative sum  $\sum_{i=0}^{t-1} I_i$  is larger or equal to the case when we already know that the sum was negative for smaller values of  $t$ .

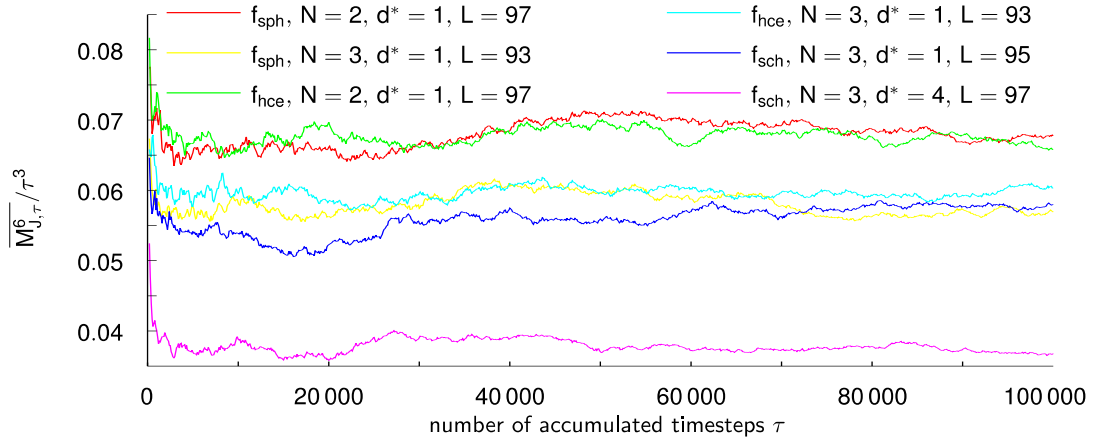
First the sixth central moment is investigated. For this purpose the previously defined estimators for the sixth central moment are visualized in Figure 12. The estimators are divided by  $\tau^3$ , because the relation to  $t^3$  or  $\tau^3$  respectively should be analyzed. With complete independence Figures 12a and 12b should appear similar to Figure 12c, which illustrates that for all scenarios the sixth central moment of the dimension dependent increment divided by  $\tau^3$  oscillates around some constant value, as it is intended to. But similar to the variance also the sixth central moment faces a heavy increase in the beginning, which results from periods with temporary additional stagnating dimensions. Nevertheless, also in Figures 12a and 12b it seems that the central sixth moment can be bounded with some polynomial of degree three, because the estimators divided by  $\tau^3$  are more or less decreasing or constant, but not increasing. Note that all theorems in this paper could be



(a) Estimator for the sixth central moment of increments



(b) Estimator for the sixth central moment of base increments



(c) Estimator for the sixth central moment of dimension dependent increments

Figure 12: Estimators for the sixth central moment of (a) increments  $\overline{M_{I,\tau}^6}$ , (b) base increments  $\overline{M_{B,\tau}^6}$  and (c) dimension dependent increments  $\overline{M_{J,\tau}^6}$  resulting from Experiment 2 divided by  $\tau^3$ ; each line represents a single configuration

adjusted if only constants  $C > 0$  and  $\varepsilon > 0$  exist, such that  $E[(\sum_{t=0}^{t-1} (I_t - \mu^*))^6] \leq C \cdot t^{4-\varepsilon}$  can be fulfilled.

The other two properties look quite reasonable.

For finite times the example runs show that there is a possibility that the increments and also the base increments and the dimension dependent increments sum up to negative values. Animated by the experimental runs, a precise PSO run can be constructed such that the increments, the base increments and the dimension dependent increments sum up to negative values. This precise run can be extended such that for each position some noise is possible, i.e., it is permitted that each reached position is at most  $\varepsilon$  apart from the intended position.  $\varepsilon$  can be chosen that small so that decisions whether a new position is a local or global optimum do not change. The probability to receive one of those runs is then positive for any finite time.

Also the property, that the probability to receive a negative sum is greater or equal, if we already know that the sum was negative before, is reasonable for the PSO. That is because mostly the previous sums are even much smaller than zero and therefore the probability to stay negative, when the next increment is added, is almost one, while it is not that likely to be negative if the previous sum is already positive.

Therefore the assumption of independence was introduced, to generate provable theorems.

#### 4.5 Brownian Motion as Approximation for Accumulated Sums

After Theorem 1 it is proposed that Brownian Motions can be used as approximations on partial sums of increments. In this section measured probabilities are compared with probabilities received from the approximation with Brownian Motion.

The probability that a stagnation phase does not end is determined by the probability that sums of increments of initial stagnating dimension stay below some bound. Actual maximal values cannot be measured, because we cannot run experiments for an infinite number of iterations. Therefore Experiment 2 is used for this purpose. For each initial stagnating dimension the maximal sum of the increments is calculated by

$$I_{\max,d} := \max_{T_m \leq t \leq T_e} \sum_{t'=T_m}^{t-1} I_{t',d}$$

and  $\langle I_{\max,d} \rangle_r$  is written for the evaluation on the  $r$ 'th run. For sure 100 000 iterations are far from infinite iterations, but on the one hand an infinite number of iterations cannot be simulated and on the other hand, as we have negative expectation of the increments, the largest values of the sums most likely appear in the beginning of the process. For the approximation with Brownian Motions the expectation and the variance of the increments are needed. For this purpose the already defined estimators  $\overline{\mu_L}$  and  $\overline{\sigma_{I,\tau}^2}$  are used. As the variance does not grow linearly from the beginning, the following bounds for the variance are used:

$$\overline{\sigma_{\max}^2} := \max_{1 \leq \tau \leq 1000} \overline{\sigma_{I,100 \cdot \tau}^2} / (100 \cdot \tau)$$

$$\overline{\sigma_{\min}^2} := \left( \overline{\sigma_{I,100000}^2} - \overline{\sigma_{I,10000}^2} \right) / 90000$$

The upper bound  $\overline{\sigma_{\max}^2}$  is the maximal measured variance per iteration. The variances are calculated only each 100'th step. The lower bound  $\overline{\sigma_{\min}^2}$  represents the approximately linear increase, neglecting the heavy increase in the beginning. After the proof of Theorem 1, it is proposed that

$1 - \exp(2(c_s - c_0)\mu/\sigma^2)$  is an approximation on the probability that the maximal sum of a single dimension stays below  $c_0 - c_s$ . Accordingly,  $1 - \exp(-2 \cdot x \cdot \mu/\sigma^2)$  can be assumed to be an approximation on the cumulative distribution function for  $I_{\max, d}$ . In Figure 13 the cumulative distribution functions, generated by the measured values  $\overline{\mu_L}$ ,  $\overline{\sigma_{\max}^2}$  and  $\overline{\sigma_{\min}^2}$ , i. e.,

$$F_{\sigma_{\max}^2}(x) := 1 - \exp(2 \cdot x \cdot \overline{\mu_L} / \overline{\sigma_{\max}^2})$$

and

$$F_{\sigma_{\min}^2}(x) := 1 - \exp(2 \cdot x \cdot \overline{\mu_L} / \overline{\sigma_{\min}^2}),$$

and the empirical cumulative distribution function, i. e.,

$$F_{\text{emp}}(x) := \frac{1}{R \cdot |D_S|} \sum_{r=1}^R \sum_{d \in D_S} \mathbb{1}_{\langle I_{\max, d} \rangle_r \leq x},$$

are visualized for two different scenarios. The two scenarios are visualized exemplarily, because

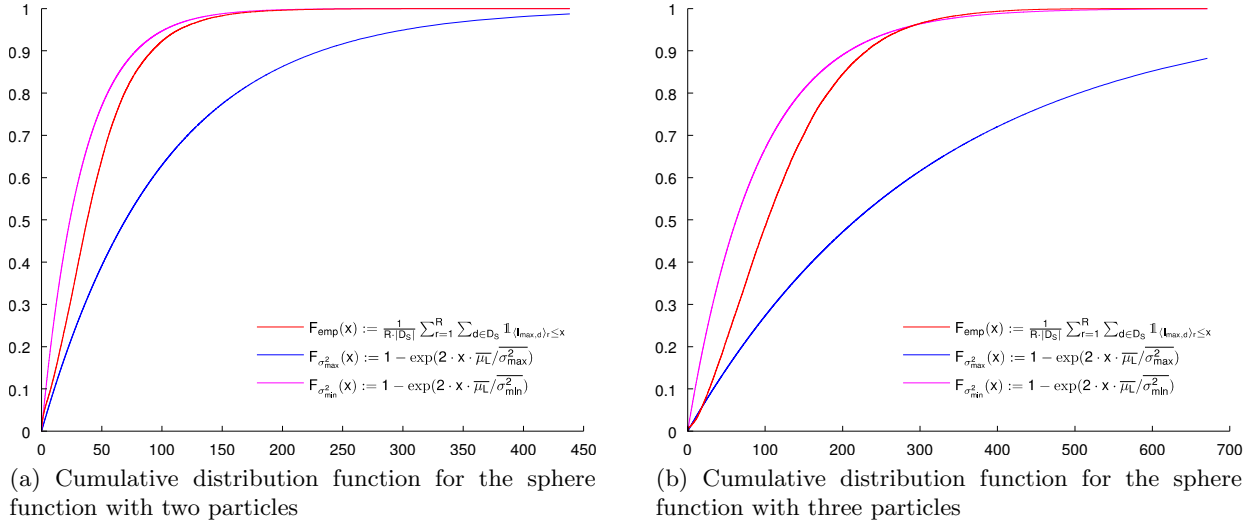


Figure 13: Cumulative distribution functions, generated by the measured values  $\overline{\mu_L}$ ,  $\overline{\sigma_{\max}^2}$  and  $\overline{\sigma_{\min}^2}$ , and the empirical cumulative distribution function for (a) the sphere function with two particles and (b) the sphere function with three particles.

for other scenarios, the respective graphs look very similar. The empirical cumulative distribution functions mainly proceed between the two cumulative distribution functions generated by the measured expectations and variances. There is some positive probability, that  $I_{\max, d}$  is 0. For the visualized scenarios this probability is quite small, as it is hardly recognizable that the empirical cumulative distribution function starts with some positive value. For small positive values it may appear that the empirical cumulative distribution functions are evaluated to lower values than the lower cumulative distribution function  $F_{\sigma_{\max}^2}$ . Furthermore, the more the value of the cumulative distribution functions tend to 1, the more likely the empirical cumulative distribution functions are evaluated to larger values than the larger cumulative function  $F_{\sigma_{\min}^2}$ . Both events occur in Figure 13b. Nevertheless, Theorem 1 only supplies positive values for the probability that  $I_{\max, d}$



stays small enough so that stagnation phases do not end, but this approximations even supply probabilities, which tend to 1 if the value for the maximal sum is enlarged. The allowed value for the maximal sum in the model used in this paper is  $c_s - c_0$ , because sums after the beginning of a stagnation phase need to stay below this value. The larger this difference is chosen, the larger is the probability that a stagnation phase never ends, at least for stagnation phases with one stagnating dimension. As more stagnating dimensions are connected with the base increment, it is more likely that the value for the maximal sum stays below some bound, if it is already known that for other dimensions this maximal sum already stays below the same bound. Therefore it is assumed that  $(1 - \exp(2(c_s - c_0)\mu/\sigma^2))^{N_0}$  is a lower bound for the probability that the maximal sum of any stagnating dimension stays below  $c_s - c_0$ , which implies that a stagnation phase does not end. The more stagnation phases occur during a single run, the more likely some of the stagnating dimensions, which do not cause the end of stagnation phases, have very small logarithmic potential. Those dimensions mainly can be neglected for calculations on the probability whether a stagnation phase ends or not. Therefore the more stagnation phases occur, the more likely those phases remain through infinity. The probability increases to values of the case with a single stagnating dimension if  $N_0 - 1$  dimensions already are that insignificant.

#### 4.6 Duration of Phases of Type $PH_X$

In this section it is discussed, why the requirement of Theorem 2, that phases of type  $PH_X$  have finite expectation, is acceptable. Afterwards distributions of durations of phases of type  $PH_X$  are visualized.

In general the event that no stagnation phase is active means that at least the largest  $D - N_0$  dimensions have quite similar potential, i. e., the absolute difference of the logarithm of their potentials is at most  $|c_0| = -c_0$ . Through the movement equations it is almost always possible, that the absolute value of the velocity in most dimensions can be decreased significantly, which leads to a significantly decreasing potential. This is even possible for  $D - 1$  dimensions. Therefore it is reasonable, that there exist some positive probability  $p$  and a finite number of iterations  $T$ , such that the probability that the logarithmic potentials of at least  $N_0$  dimensions decrease to values less than  $c_0$  within  $T$  iterations is at least  $p$ . Then the waiting time for the event that the logarithmic potential of at least  $N_0$  dimensions decrease to values less than  $c_0$  is bounded above by a geometrical distribution, which has the finite expectation  $T/p$ . Therefore the expectation of the waiting time for the next stagnation phase is also bounded by  $T/p$ , but this stagnation phase do not necessarily achieve the limits  $\mu$ ,  $M$  or  $p_0$  specified in Definition 6. The choices of the values  $N_0$ ,  $\mu$ ,  $M$ , and  $p_0$  are heavily dependent on the scenario.

For instance the sphere function  $f_{sph}$  or the high conditioned elliptic function  $f_{hce}$  with two particles can be analyzed. As already discussed, it only matters how many dimensions are stagnating. Therefore all stagnation phases are good stagnation phases with equal parameters  $\mu$ ,  $M$ , and  $p_0$  if  $N_0 = D - 3$  dimensions are stagnating. If more dimensions than  $N_0$  are stagnating, then the expectation of the increments is positive, which finally lead to fewer number of stagnating dimensions. This leads to the appraisal that  $T/p$  is also a bound for the expected waiting time to the next good stagnation phase for this scenario.

In contrast, the Schwefel's problem  $f_{sch}$  is not that easy to handle, because the behavior of the increments depends on the set of stagnating dimensions. To guarantee suitable values for  $N_0$ ,  $\mu$ ,  $M$  and  $p_0$ , most subsets of the set of dimensions need to be analyzed, whether this set may appear as stagnating dimensions or not. As there are  $2^D$  many subsets, this cannot be analyzed

with acceptable time required for computing if  $D$  becomes large. Nevertheless  $N_0$  can be chosen a little bit lower so that additional, not only temporary, stagnating dimensions are allowed. Without knowing the exact values,  $\mu$  and  $M$  can be chosen as the maximal value over all good stagnation phases and  $p_0$  can be chosen as the minimal value over all good stagnation phases. Also the experiments show that stagnation appears for Schwefel's problem as well.

For all analyzed scenarios the parameters  $N_0$ ,  $c_0$ ,  $c_s$ ,  $\mu$ ,  $M$  and  $p_0$  can be chosen, such that all observed stagnation phases are good stagnation phases, because the smaller  $c_0$  is, i. e., the larger the absolute value of  $c_0$  is, the more unlikely are stagnation phases with positive expectation of increments. In particular, this implies  $\alpha_i = \tilde{\alpha}_i$  and  $\beta_i = \tilde{\beta}_i$  for every  $i$ , which is assumed from now on.

The following experiment supplies data for actual stagnation phases, because a usual initialization is used.

**Experiment 3.** *All functions are evaluated with 500 000 iterations. The number of particles  $N$ , the number of dimensions  $D$  and the objective functions vary. Usual initialization is used, i. e., Algorithm 2 can be used for position initialization, but  $L$ , the number of initially stagnating dimensions, is set to zero, and all velocities are initially set to zero. Each tested configuration is started  $R = 500$  times with different seeds. Let  $T_e$  be 500 000, the number of time steps of the evaluated process.  $\Delta t$  is set to 100.*

The following abbreviations are defined:

$$D_{X,i} := \{r \in \{1, \dots, R\} : i = 0 \vee \langle \beta_{i-1} \rangle_r \leq T_e\}$$

represents the set of runs such that the  $i$ 'th phase of type  $PH_X$  has started within the tested number of iterations.

$$D_{Y,i} := \{r \in \{1, \dots, R\} : \langle \alpha_i \rangle_r \leq T_e\}$$

represents the set of runs such that the  $i$ 'th phase of type  $PH_Y$  or  $PH_F$  has started within the tested number of iterations. For all tested scenarios  $D_{X,i}$  equals  $D_{Y,i}$ , which means that no test run of any tested scenario ended in a phase of type  $PH_X$ .  $|D_{X,0}| = R = 500$ , because each test run starts with a phase of type  $PH_X$ . Furthermore,

$$F_{X,i}(x) := \sum_{r \in D_{X,i}} \frac{\mathbb{1}_{\langle X_i \rangle_r \leq x}}{|D_{X,i}|}$$

represents the empirical distribution of the phase lengths  $X_i$ . Figure 14 shows this empirical distribution, received from Experiment 3, for the durations of a phase of type  $PH_X$  with the sphere function  $f_{sph}$ ,  $N = 3$  particles,  $D = 100$  dimensions,  $N_0 = D - 7 = 93$ ,  $c_0 = -40$  and  $c_s = -20$ . This configuration is chosen exemplarily. For all configurations the durations of a phase of type  $PH_X$  are largest for the initial phase of type  $PH_X$ , which starts at iteration 0. At the beginning, all dimensions have quite the same potential. Therefore the logarithmic potential of all dimensions are not that far from zero. This implies that  $N_0$  dimensions need to decrease their potential to  $c_0 = -40$  or lower. After the end of a stagnation phase it is only known, that from the set of initially stagnating dimensions less than  $N_0$  dimensions remain, which have logarithmic potential less than  $c_s = -20$  since the start of the stagnation phase. Nevertheless there can be either other dimensions, which now have already low logarithmic potential, i. e., less than  $c_0$  or  $c_s$ . Also there might be dimensions, which did not cause the end of the stagnation phase and still have

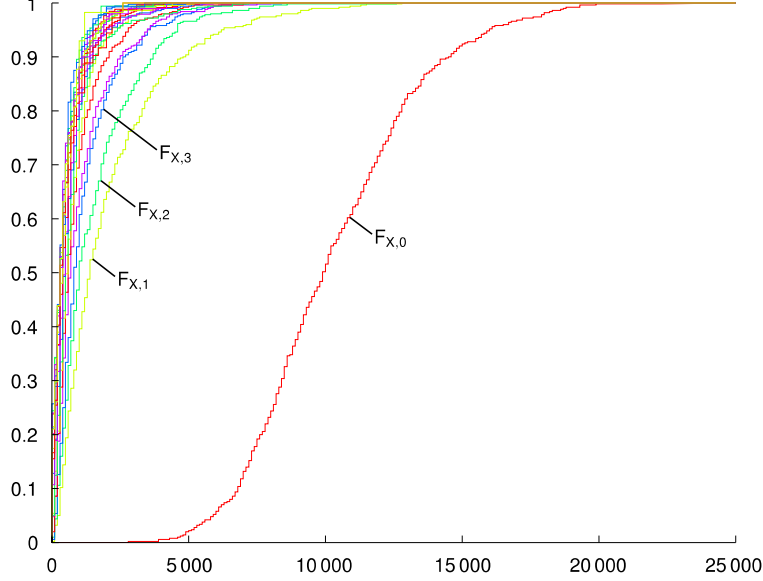


Figure 14: Empirical distribution of phase lengths  $X_i$  with sphere function,  $N = 3$  particles,  $D = 100$  dimensions,  $N_0 = 93$ ,  $c_0 = -40$  and  $c_s = -20$

low logarithmic potential. Additionally, if the logarithmic potential of a dimension is larger than  $c_s$ , then that logarithmic potential is not almost zero. Instead it most likely is just a little bit larger than  $c_s$ . Altogether, after stagnation phases there are still dimensions with much lower logarithmic potential than they most likely have immediately after the initialization of the process. Therefore the data is conform with this theoretical thoughts. For this large number of stagnating dimensions  $N_0 = 97$  it is also reasonable, that the graphs of the empirical distributions move even further to the left, because it becomes more and more likely that some of the stagnating dimensions are much smaller than  $c_0$  and therefore do not delay the start of the next stagnation phase. Further data on the existence of phases of type  $PH_X$  and some statistical values can be found in Appendix B.

#### 4.7 Swarm Converges

In this section it is discussed why the swarm converges almost surely for the tested scenarios.

On the one hand the chosen parameters  $\chi = 0.72984$  and  $c_1 = c_2 = 1.496172$ , as proposed in [4], are known to be good, because rough theoretical analysis shows, that convergence can be expected.

On the other hand the experiments supply also information, which lead to the assumption of convergence. As assumed, the logarithmic potential of stagnating dimensions tends to minus infinity with linear drift. Also the logarithm of the potential of all dimensions tend to minus infinity with linear drift for all recently mentioned scenarios. The derivative for a stagnating dimension  $d_1$  finally stays quite constant. Therefore the potential in dimension  $d_1$  is approximately the absolute value of the velocity in that dimension multiplied by that constant:

$$\Phi(t, d_1) \approx \max_{n \in \{1, \dots, N\}} |V_t^{n, d_1}| \cdot \text{const.}$$

The logarithm of the potential is then the logarithm of the absolute value of the velocity plus the

logarithm of a constant value:

$$\log(\Phi(t, d_1)) \approx \log \left( \max_{n \in \{1, \dots, N\}} |V_t^{n, d_1}| \right) + \log(const)$$

This implies that also the logarithm of the absolute values of the velocities of stagnating dimensions tends to minus infinity with linear drift. For a non-stagnating dimension  $d_2$  the first derivative is approximately zero, but the second derivatives of all used functions are positive definite matrices. Therefore it can be assumed that the potential of non-stagnating dimensions can be approximated by the square of the velocity multiplied by a constant:

$$\Phi(t, d_2) \approx \max_{n \in \{1, \dots, N\}} (V_t^{n, d_2})^2 \cdot const.$$

The logarithm of the potential is then two times the logarithm of the absolute value of the velocity plus the logarithm of a constant value:

$$\log(\Phi(t, d_2)) \approx 2 \cdot \log \left( \max_{n \in \{1, \dots, N\}} |V_t^{n, d_2}| \right) + \log(const)$$

This implies that also the logarithm of the absolute value of the velocities of non-stagnating dimensions tend to minus infinity with linear drift.

Therefore the absolute values of the velocities  $V_t^{d, n}$  can be approximately bounded by  $v_0 \cdot c^t$  with positive constants  $v_0 \in \mathbb{R}$  and  $c \in ]0, 1[$ . The maximal distance, which can be covered by a particle beginning at some time  $T$ , is then bounded by

$$\sum_{t=T}^{\infty} v_0 \cdot c^t = v_0 \cdot c^T \sum_{t=0}^{\infty} c^t = \frac{v_0 \cdot c^T}{1 - c}.$$

This bound is a finite value which tends to zero if  $T$  tends to infinity. Therefore the positions of the particles will tend to constant positions. This position is indeed the limit of the global attractor, because otherwise the velocities of a particle will stay approximately as large as the difference from current position to global attractor and will not tend to zero.

Another point of view is that all investigated functions are strictly convex, i. e.,

$$f(a \cdot x + (1 - a)y) < a \cdot f(x) + (1 - a)f(y)$$

for all  $a \in ]0, 1[$  and  $x, y \in \mathbb{R}^D$ ,  $x \neq y$ . Usually the positions are initialized in a closed set. Therefore the maximal function value for the initial global attractor  $f(G_1^1)$  is bounded. As all investigated functions can be written as  $x^t \cdot A \cdot x$ , with positive definite matrix  $A$ , the set of points with lower function values than the worst possible initial global attractor is bounded too. With the theorem of Bolzano-Weierstraß we receive that  $(G_t^1)_{t \in \mathbb{N}_0}$  contains convergent subsequences. If there is more than one accumulation point, then the function value of all accumulation points needs to be equal, because the functions are continuous and therefore points with a larger function value than the lowest function value of any accumulation point cannot remain as accumulation points. But if there are different accumulation points, then the particles almost surely visit the area between the accumulation points. As the investigated functions are strictly convex, the points between two points with equal function value have strictly lower function values than the function values of the accumulation points. After one of those points in between is visited, the global attractor will

change to that point and will never return to the surroundings of the intended accumulation points, because the function values in this area are larger than the encountered value of the current point. Therefore the sequence of the global attractor can have at most one accumulation point almost surely. The theorem of Bolzano-Weierstraß guarantees, that there is at least one accumulation point. This implies that there is exactly one accumulation point  $G_{lim}$  almost surely. If the global attractor does not converge to that accumulation point, then there exists an  $\varepsilon > 0$  such that the set  $I := \{t \in \mathbb{N}_0 : \|G_t^1 - G_{lim}\|\}$  does not have a finite number of elements. With the theorem of Bolzano-Weierstraß also a subsequence of the sequence only containing indexes available in  $I$  has at least one accumulation point. This accumulation point cannot be equal to  $G_{lim}$ , but this is a contradiction to the statement that there is only a single accumulation point. Therefore the complete sequence of the global attractors converges almost surely to the limit point  $G_{lim}$ . Indeed that limit point is not deterministic. Also the local attractors need to converge to the same limit point almost surely, because otherwise the same contradictions appear as with multiple accumulation points of the global attractor. With usual parameter choices then also the positions will converge to the limit point of the global attractor and the velocities will tend to zero.

Nevertheless the property, that a swarm will converge almost surely, is dependent on the objective function, the number of particles, the number of dimensions and the parameters of the PSO and needs to be checked if Theorem 2 should be applied.

#### 4.8 Empirical Distributions of Stopping Times $\alpha_i$ and $\beta_i$ and of Logarithmic Potential

This section provides additional data concerning Experiment 3. First, empirical distributions of the stopping times  $\alpha_i$  and  $\beta_i$  are provided for two different scenarios. Afterwards, empirical distributions of the logarithmic potential at the end of the test runs are visualized.

The empirical distributions of the stopping times  $\alpha_i$  and  $\beta_i$  are defined by the following expressions:

$$F_{\alpha,i}(x) := \frac{1}{R} \sum_{r=1}^R \mathbb{1}_{\langle \alpha_i \rangle_r \leq x}, \quad F_{\beta,i}(x) := \frac{1}{R} \sum_{r=1}^R \mathbb{1}_{\langle \beta_i \rangle_r \leq x}$$

Figure 15 shows the empirical distributions for the configuration with sphere function,  $N = 3$  particles,  $D = 100$  dimensions,  $N_0 = 93$ ,  $c_0 = -40$  and  $c_s = -20$  and Figure 16 shows the empirical distributions for the configuration with sphere function,  $N = 3$  particles,  $D = 8$  dimensions,  $N_0 = 1$ ,  $c_0 = -40$  and  $c_s = -20$ . The duration of the first phase of type  $PH_X$  takes longer with more dimensions. Also it is more unlikely that stagnation phases remain. Especially in case of  $D = 100$  dimensions in the beginning most stagnation phases are quite short, because in 93 dimensions the logarithmic potential needs to stay small. For  $D = 8$  only the logarithmic potential of a single dimension needs to stay small and therefore the probability that a stagnation phase remains is much higher. For  $D = 100$  it becomes also more and more likely that stagnation phases remain, because there are more and more dimensions, which have a logarithmic potential which remains far below  $c_0$ . This can also be observed in the Figures 15 and 16. The leftmost line represents the empirical distribution of  $\alpha_0$ , which equals the length of the first phase of type  $PH_X$ . For  $D = 100$  the first empirical distributions of the stopping times  $\alpha_i$  and  $\beta_i$  reach the value 1 very early. For  $D = 8$  the empirical distribution of  $\beta_0$ , which is represented by the second left line, indicates the end of the first stagnation phase. In contrast to the previous scenario, this empirical distribution does not reach the value 1. That means that the first stagnation phase does not end within the

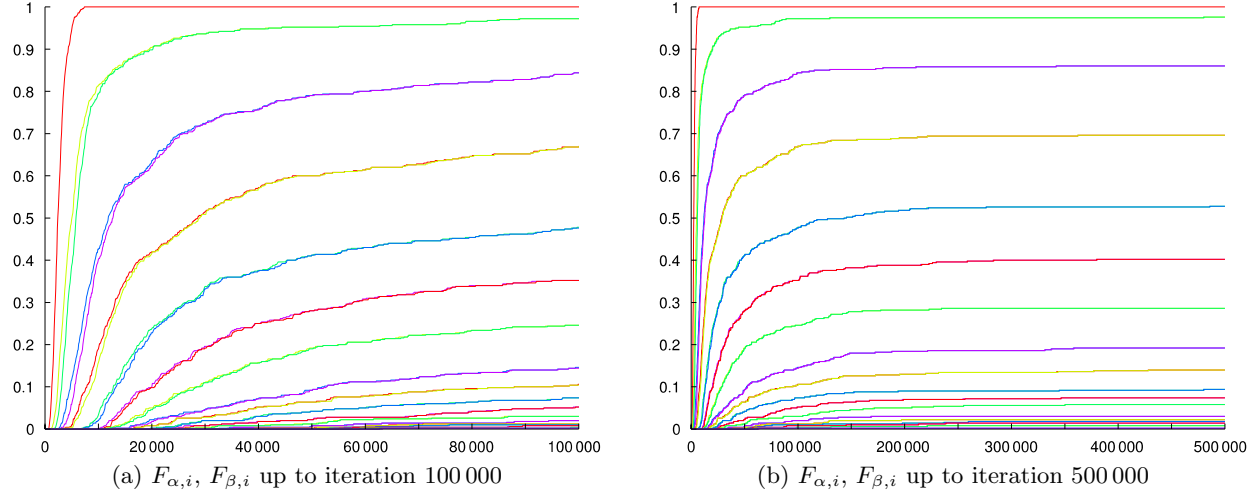


Figure 15: Empirical distribution of the stopping times  $\alpha_i$  and  $\beta_i$  ( $F_{\alpha,i}, F_{\beta,i}$ ) with sphere function,  $N = 3$  particles,  $D = 8$  dimensions,  $N_0 = 1$ ,  $c_0 = -40$  and  $c_s = -20$

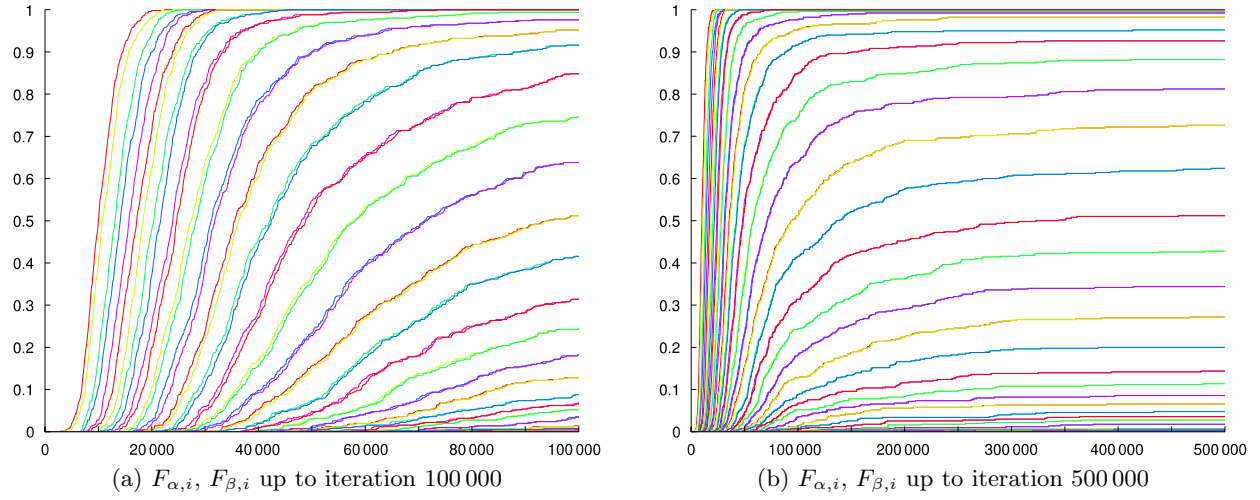


Figure 16: Empirical distribution of the stopping times  $\alpha_i$  and  $\beta_i$  ( $F_{\alpha,i}, F_{\beta,i}$ ) with sphere function,  $N = 3$  particles,  $D = 100$  dimensions,  $N_0 = 93$ ,  $c_0 = -40$  and  $c_s = -20$

tested iterations. A further observation is that phases of type  $PH_X$  have quite short duration compared with phases of type  $PH_Y$ . Already in Figures 15a and 16a the empirical distributions of  $F_{\beta,i}$  and  $F_{\alpha,i+1}$  have only a small distance. In Figures 15b and 16b they are merging to a single line already.

Another indication for the existence of unlimited stagnation phases is the empirical distribution of the final value of the logarithmic potential.

$$F_{\Psi,d}(x) := \frac{1}{R} \sum_{r=1}^R \mathbb{1}_{(\sum_{d'=1}^D \mathbb{1}_{\langle \Psi(T_e, d') \rangle_r \leq x}) \geq d}$$

To receive more meaningful figures, not the empirical distributions for single dimensions are visualized, but the empirical distributions for the  $d$ 'th largest dimension is visualized.  $\sum_{d'=1}^D \mathbb{1}_{\langle \Psi(T_e, d') \rangle_r \leq x}$  counts how many dimensions of run  $r$  have a final logarithmic potential of less or equal than  $x$ . A single run has then only an effect if at least  $d$  dimensions with final logarithmic potential of less or equal than  $x$  appear. Figures 17 and 18 show the respective empirical distributions for two scenarios.

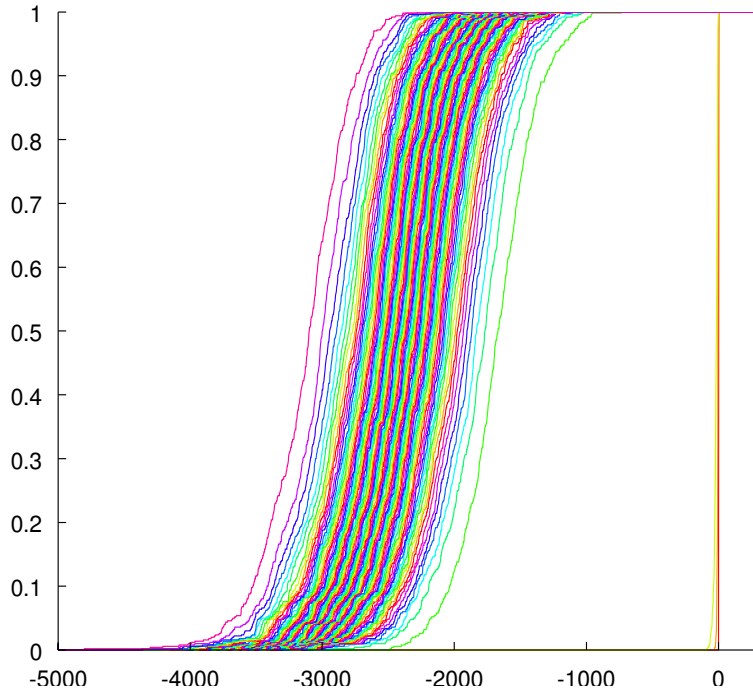


Figure 17: Empirical distribution of the final value of the logarithmic potential  $F_{\Psi,d}$  with sphere function,  $N = 2$  particles and  $D = 100$  dimensions

As already proposed, the scenario with the sphere function and two particles, which is displayed in Figure 17, experiences always  $D - 3$  stagnating dimensions. The three dimensions with largest logarithmic values at the end significantly differ from the other dimensions. Their logarithmic potential is not that far from zero. In contrast the logarithmic potential of the remaining 97 dimensions is always smaller than  $-500$  and in most cases much smaller than this value. This small value means that the potential of the fourth largest dimension is  $2^{-500} \approx 10^{-151}$  times smaller than the potential of the largest dimension. The minimal logarithmic potential value of the third

largest dimension is approximately  $-90$  and the minimal logarithmic potential value of the second largest dimension is approximately  $-30$ . Admittedly, these values are also very small compared to the largest dimension, but, as already discussed, dimensions can become temporarily stagnating. Therefore small logarithmic potential values can appear even for the dimension with second largest potential. Nevertheless, the difference from the third to the fourth largest logarithmic potential value is huge. In contrast the empirical distributions of the dimensions with lower logarithmic potential are quite similar. They only move a little bit to the left.

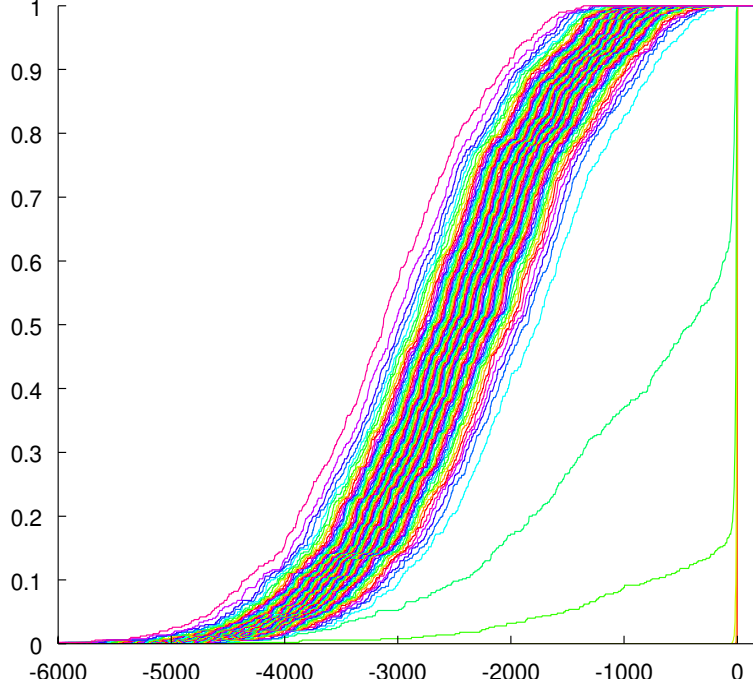


Figure 18: Empirical distribution of the final value of the logarithmic potential  $F_{\Psi,d}$  with Schwefel's problem,  $N = 3$  particles and  $D = 100$  dimensions

For the Schwefel's problem the number of actual stagnating dimensions depends on which dimensions are stagnating and which are not. There are sets with three non-stagnating dimensions such that the expectation of the increments becomes negative, but there are also sets with five non-stagnating dimension such that, if any of the dimensions is removed from the set of non-stagnating dimensions, then the expectation of the increments becomes positive. Figure 18 represents the empirical distributions of the Schwefel's problem with three particles and 100 dimensions. There it can be observed that the empirical distribution of the fourth and fifth largest logarithmic potential value do neither always belong to the set of stagnating dimensions nor they always belong to the set of non-stagnating dimensions. Instead there is some positive probability that the dimension, which refers to the fourth or the fifth largest logarithmic potential value, belongs to the set of non-stagnating dimension. This is illustrated by the heavy increase of the empirical distribution in the near surroundings of the potential value zero. For the remaining cases the fourth or the fifth largest logarithmic potential value behave similar to the other stagnating dimensions. It is also reasonable that the fourth or the fifth largest logarithmic potential value behave not identical, because the distribution of the increments depends on the set of stagnating dimensions. For sixth



and further dimensions all stagnation phases are merged, while the fifth largest potential only counts as stagnating dimension if four or fewer dimensions are non-stagnating. The cases with exactly five non-stagnating dimensions is removed for this dimension.

## 4.9 Long Time Experiments

Further more time consuming experiments are done with larger number of particles. Beware, if the number of steps is doubled, then the runtime sometimes increases by a factor of four, because the longer the experiment takes the more increases the precision. This is due to the event that the longer an experiment takes the more diverse become the values. Also the number of dimensions influences the runtime more than linearly. To grant no irregularities the precision needs to be increased, which is executed automatically on demand by the used software. As this experiments are analyzed for 100 000 000 iterations only single runs are executed.

For example the configuration with four particles and the sphere function was under investigation. With  $D = 40$  dimensions and 100 000 000 iterations it appears that after a long period of time the smallest occurring logarithmic potential among the dimensions stays quite constant. Indeed the dimension, which is responsible for the lowest value of the logarithmic potential changes over time. The value for the lowest logarithmic potential  $\Psi(t, d)$  stays approximately in the range  $-1\,000$  to  $-2\,000$ . Periods of time with temporary stagnating dimensions can even last up to 10 000 000 iterations. However the largest overall potential  $\Phi(100\,000\,000, d)$  has decreased to approximately  $2^{-8\,000}$ . It even appears that each dimension becomes significant from time to time, which leads to the impression that four particles finally will succeed to optimize the sphere function with 40 dimensions. The absolute values of the global attractor for all dimensions at the end are in the range from  $2^{-4\,200}$  to  $2^{-3\,600}$ , which strengthens this impression.

With  $D = 400$  dimensions and 100 000 000 iterations it cannot yet be determined, whether there are actual stagnating dimensions or not. In fact the value for the lowest logarithmic potential  $\Psi(t, d)$  reaches approximately  $-10\,000$ , but no significant separation of logarithmic potential of different dimensions can be observed, which would indicate stagnating and non-stagnating dimensions. Some dimensions are stagnating almost since the beginning of that test run and have very small logarithmic potential, but other dimensions regained logarithmic potential after very large number of iterations. The maximal potential value of the last iterations has not decreased that extremely. Potential values  $\Phi(t, d)$  of up to approximately  $2^{-30}$  are reached in the last 10% of the measured iterations. The absolute values of the global attractor for all dimensions at the end are in the range from  $2^{-250}$  to  $2^7$ . This means that the position of the global attractor in some dimensions at the end is still in the range of the initialized positions. Surely this looks like that there are permanent stagnating dimensions, but, as periods of time with temporary stagnating dimensions can take even more iterations than in the case with 40 dimensions, no prediction can be made, whether there are stagnating dimensions in this case or not. On the one hand a larger number of iterations needs to be tested and on the other hand more than a single test run needs to be executed. This is not quite easy because already this single test run took some months of computing time, because the final needed precision is more than 10 000 bits, which cause very time consuming calculations.

Therefore it is not quite clear, whether there is a limit of dimensions such that the PSO with chosen parameters and four particles finally stagnates at non-optimal points, or not.

## 5 Conclusion

Altogether a theoretical base was introduced to formalize stagnation during runs of the PSO. For some functions evidence was provided, that the chosen theoretical model is applicable. Especially for the sphere function and its scaled version, the high conditioned elliptic function, it was shown that with two particles there are at most three dimensions, which do not stagnate permanently. All other dimensions are finally stagnating. For three particles the number of dimensions, which do not stagnate permanently, is increased to seven. For three particles, and PSO parameters as used in this paper, the sphere function and eight dimensions there is always a dimension, which finally stops optimizing. For other functions the number of stagnating dimensions can be significantly different from the configuration with the sphere function. An example is the diagonal function. Even with four particles there are only two dimensions, which do not stagnate permanently.

Furthermore, the presented framework is not only applicable to the presented version of the PSO. Everything can work also with different parameters for the movement equations or even with changed rules for movement or updates. For example, the global attractor could be updated only each iteration and not after each particle. Also other iteration based optimization algorithms can be analyzed if a suitable potential can be defined.

On the one hand it can be analyzed whether optimization algorithms can find local optima for specified functions, or it may appear that there are always some dimensions, such that the coordinates in this dimension are not optimized. On the other hand if convergence is present and if comparable definitions for a potential can be made, for example Definition 3 for PSO algorithms, then the speed of convergence can be compared by comparison of the increments of logarithmic potentials. Also the number of stagnating dimensions can be a comparison criterion.

In this paper only functions are analyzed, which do not change their shape, no matter how the coordinates are scaled. For example if there are two vectors  $x, y \in \mathbb{R}^D$  then  $f(a \cdot x) \leq f(a \cdot y) \Leftrightarrow a^2 f(x) \leq a^2 f(y) \Leftrightarrow f(x) \leq f(y)$  for any  $a \in \mathbb{R} \setminus \{0\}$ . The established framework can also handle more complicated functions, for instance functions with bounded third derivatives. For each local optimum those functions can be approximated by their Taylor approximation of degree two. If the second derivative matrix is not only positive semidefinite but also positive definite, then this approximation is sufficient to determine whether the PSO finally stops to tend to that local optimum or not. If the PSO running on the approximation almost surely stagnates and does not end up in the single optimum, then the PSO will stagnate if it tends to that local optimum in the original function, because changes in higher derivatives become less important the nearer the particles are positioned around the local optimum. In contrast if for all local optima stagnation is impossible almost surely for the Taylor approximation, then it can also stagnate before, when higher derivatives are still significant.

Additionally if stagnation is not present then convergence to local optima is not granted. At least not with the recent analysis. Nevertheless it is expected that if stagnation is not available, then each dimension reaches large logarithmic potential from time to time. If dimensions have large logarithmic potential, then it is assumed that these dimensions optimize their positions. In this paper stagnation is defined for some minimal number of stagnating dimensions  $N_0$ , which is appropriate to grant not reaching a local optimum. To grant convergence to local optima for at least  $D - N_0$  dimensions, stagnation phases need to be defined differently with a maximal (not minimal) number of stagnating dimensions. This positive result can not be extended by only analyzing the derivatives of local optima, because earlier stagnation can be available.

Nevertheless if there is no positive value of  $N_0$ , the minimal number of stagnating dimensions, such that stagnation is observed, then it is expected that convergence to local optima is present.

Final continuing questions in this context, which could be analyzed with the given framework, are the following: Is there a number of particles, such that it can be granted that at least the sphere function, continuous functions or differentiable functions will be optimized perfectly almost surely? Is there a fixed number of particles such that it can be granted that at least the sphere function, continuous functions or differentiable functions will be optimized perfectly almost surely for some bounded number of dimensions? Is there a minimal or maximal number of stagnating dimensions for fixed number of particles, fixed PSO parameters and a set of objective functions?

## References

- [1] H. Bauer. *Probability Theory*, volume 23 of *Studies in Mathematics*. De Gruyter, 1996.
- [2] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [3] A. Carlisle and G. Dozier. An off-the-shelf PSO. In *Proc. Particle Swarm Optimization Workshop*, pages 1–6, 2001.
- [4] M. Clerc and J. Kennedy. The particle swarm – explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 6:58–73, 2002.
- [5] R. Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2010.
- [6] R. C. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In *Proc. 6th International Symposium on Micro Machine and Human Science*, pages 39–43, 1995.
- [7] M. Jiang, Y. P. Luo, and S. Y. Yang. Particle swarm optimization – stochastic trajectory analysis and parameter selection. In F. T. S. Chan and M. K. Tiwari, editors, *Swarm Intelligence – Focus on Ant and Particle Swarm Optimization*, chapter 17, pages 179–198. I-TECH Education and Publishing, Vienna, 2007.
- [8] M. Jiang, Y. P. Luo, and S. Y. Yang. Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm. *Inf. Process. Lett.*, 102:8–16, 2007.
- [9] J. Kennedy and R. C. Eberhart. Particle swarm optimization. In *Proc. IEEE International Conference on Neural Networks*, volume 4, pages 1942–1948, 1995.
- [10] P. K. Lehre and C. Witt. Finite first hitting time versus stochastic convergence in particle swarm optimisation. In L. Di Gaspero, A. Schaerf, and T. Stützle, editors, *Advances in Metaheuristics*, OR/CS, pages 1–20. Springer Science+Business Media, New York, USA, 2013.
- [11] J. E. Onwunali and L. J. Durlofsky. Application of a particle swarm optimization algorithm for determining optimum well location and type. *Computational Geosciences*, 14:183–198, 2010.
- [12] B. K. Panigrahi, Y. Shi, and M.-H. Lim, editors. *Handbook of Swarm Intelligence — Concepts, Principles and Applications*. Springer, 2011.
- [13] K. Ramanathan, V. M. Periasamy, M. Pushpavanam, and U. Natarajan. Particle swarm optimisation of hardness in nickel diamond electro composites. *Archives of Computational Materials Science and Surface Engineering*, 1:232–236, 2009.
- [14] T. H. Scheike. A boundary-crossing result for the Brownian Motion. *Journal of Applied Probability*, 29(2), 1992.

- [15] B. I. Schmitt. *Convergence Analysis for Particle Swarm Optimization*. FAU University Press, Erlangen, 2015. Doctoral Thesis.
- [16] M. Schmitt and R. Wanka. Particle swarm optimization almost surely finds local optima. In *Proc. 15th Genetic and Evolutionary Computation Conference (GECCO)*, pages 1629–1636, 2013.
- [17] M. Schmitt and R. Wanka. Particles prefer walking along the axes: Experimental insights into the behavior of a particle swarm. In *Companion of Proc. 15th Genetic and Evolutionary Computation Conference (GECCO)*, pages 17–18, 2013.
- [18] M. Schmitt and R. Wanka. Particle swarm optimization almost surely finds local optima. *Theoretical Computer Science*, 561:57–72, 2015.
- [19] P. N. Suganthan, N. Hansen, J. J. Liang, K. Deb, Y. P. Chen, A. Auger, and S. Tiwari. Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimization. Technical report, Nanyang Technological University, Singapore, 2005.
- [20] I. C. Trelea. The particle swarm optimization algorithm: Convergence analysis and parameter selection. *Inf. Process. Lett.*, 85:317–325, 2003.
- [21] M. P. Wachowiak, R. Smolíková, Y. Zheng, J. M. Zurada, and A. S. Elmaghraby. An approach to multimodal biomedical image registration utilizing particle swarm optimization. *IEEE Transactions on Evolutionary Computation*, 8:289–301, 2004.

## A Data of Experiment 1

Table 1: Values of estimaters specified in Experiment 1 for the sphere function  $f_{sph}$

$f$	$N$	$L$	$d^*$	$\bar{\mu}_U$	$\bar{\mu}_M$	$\bar{\mu}_D$	$\bar{\mu}_L$
$f_{sph}$	2	9	1	-0.280370	-0.280370	-0.067542	0.212828
$f_{sph}$	2	8	1	-0.133841	-0.133843	-0.0742897	0.0595512
$f_{sph}$	2	7	1	-0.0265214	-0.0265178	-0.0307508	-0.0042293
$f_{sph}$	3	9	1	-0.281658	-0.281658	-0.0653937	0.216264
$f_{sph}$	3	8	1	-0.232753	-0.232748	-0.0736407	0.159112
$f_{sph}$	3	7	1	-0.175692	-0.175695	-0.0677883	0.107903
$f_{sph}$	3	6	1	-0.0944085	-0.094405	-0.0458445	0.048564
$f_{sph}$	3	5	1	-0.0436236	-0.043621	-0.026277	0.0173466
$f_{sph}$	3	4	1	-0.0202586	-0.020237	-0.0155185	0.0047401
$f_{sph}$	3	3	1	-0.0074307	-0.0074679	-0.0091868	-0.0017561

Table 2: Values of estimaters specified in Experiment 1 for the high conditioned elliptic function  $f_{hce}$

$f$	$N$	$L$	$d^*$	$\bar{\mu}_U$	$\bar{\mu}_M$	$\bar{\mu}_D$	$\bar{\mu}_L$
$f_{hce}$	2	9	1	-0.280370	-0.280370	-0.067542	0.212828
$f_{hce}$	2	8	1	-0.133605	-0.133599	-0.0742256	0.0593796
$f_{hce}$	2	7	1	-0.026488	-0.026485	-0.0307987	-0.0043110
$f_{hce}$	3	9	1	-0.281658	-0.281658	-0.0653937	0.216264
$f_{hce}$	3	8	1	-0.232727	-0.232721	-0.073618	0.159109
$f_{hce}$	3	7	1	-0.175414	-0.175412	-0.0676028	0.107811
$f_{hce}$	3	6	1	-0.0941275	-0.0941295	-0.0457537	0.0483738
$f_{hce}$	3	5	1	-0.0436067	-0.0436075	-0.0263199	0.0172867
$f_{hce}$	3	4	1	-0.0201859	-0.0201904	-0.0154480	0.0047379
$f_{hce}$	3	3	1	-0.0074169	-0.0074541	-0.0091259	-0.0017089

Table 3: Values of estimaters specified in Experiment 1 for the Schwefel's problem  $f_{sch}$

$f$	$N$	$L$	$d^*$	$\bar{\mu}_U$	$\bar{\mu}_M$	$\bar{\mu}_D$	$\bar{\mu}_L$
$f_{sch}$	2	9	1	-0.280217	-0.280217	-0.0675834	0.212634
$f_{sch}$	2	9	2	-0.280472	-0.280472	-0.0676385	0.212833
$f_{sch}$	2	8	1	-0.0920762	-0.0920796	-0.0622003	0.0298759
$f_{sch}$	2	8	3	-0.0258683	-0.0258645	-0.0449234	-0.0190551
$f_{sch}$	2	7	1	-0.0052527	-0.0052403	-0.0215513	-0.0162987
$f_{sch}$	2	7	4	-0.0275117	-0.0446342	-0.0450705	-0.0175588
$f_{sch}$	3	9	1	-0.281479	-0.281479	-0.0654046	0.216074
$f_{sch}$	3	8	1	-0.220181	-0.220177	-0.0691141	0.151067
$f_{sch}$	3	8	3	-0.166589	-0.166591	-0.0553963	0.111193
$f_{sch}$	3	7	1	-0.120368	-0.120371	-0.0522450	0.0681226
$f_{sch}$	3	7	4	-0.0585785	-0.058586	-0.0338823	0.0246963
$f_{sch}$	3	6	1	-0.0363988	-0.036386	-0.0251476	0.0112512
$f_{sch}$	3	6	5	-0.0169291	-0.0169333	-0.0173492	-0.0004201
$f_{sch}$	3	5	1	-0.0083174	-0.0083236	-0.0120229	-0.0037055
$f_{sch}$	3	5	6	-0.0080515	-0.0086733	-0.0125910	-0.0045394
$f_{sch}$	3	4	1	-0.0076700	-0.0116463	-0.0117998	-0.0041299
$f_{sch}$	3	4	7	-0.0065029	-0.0118378	-0.0116454	-0.0051425
$f_{sch}$	3	3	1	-0.0071813	-0.0118641	-0.0115253	-0.0043441
$f_{sch}$	3	3	8	-0.0055640	-0.0114905	-0.0110172	-0.0054532

Table 4: Values of estimaters specified in Experiment 1 for the diagonal function  $f_{diag}$

$f$	$N$	$L$	$d^*$	$\bar{\mu}_U$	$\bar{\mu}_M$	$\bar{\mu}_D$	$\bar{\mu}_L$
$f_{diag}$	2	9	1	-0.280374	-0.280374	-0.067633	0.212741
$f_{diag}$	2	8	1	$-6.88 \cdot 10^{-6}$	$-6.33 \cdot 10^{-6}$	-0.037121	-0.037114
$f_{diag}$	3	9	1	-0.281716	-0.281716	-0.065535	0.216181
$f_{diag}$	3	8	1	$-1.09 \cdot 10^{-5}$	$-1.08 \cdot 10^{-5}$	-0.027709	-0.027699
$f_{diag}$	4	9	1	-0.281934	-0.281934	-0.064976	0.216958
$f_{diag}$	4	8	1	$-1.58 \cdot 10^{-5}$	$-1.83 \cdot 10^{-5}$	-0.006380	-0.006364

## B Data of Experiment 3

The following variables are defined, which indicate minimal, maximal and average values and variance observed with Experiment 3:

$$\min_{X,i} := \inf_{r \in D_{X,i}} \langle X_i \rangle_r, \quad \max_{X,i} := \sup_{r \in D_{X,i}} \langle X_i \rangle_r,$$

$$\mu_{X,i} := \sum_{r \in D_{X,i}} \frac{\langle X_i \rangle_r}{|D_{X,i}|}, \quad \text{and} \quad \sigma_{X,i}^2 := \sum_{r \in D_{X,i}} \frac{(\langle X_i \rangle_r - \mu_{X,i})^2}{|D_{X,i}|}.$$

Table 5: Evaluation of data from Experiment 3 with  $c_0 = -40$ ,  $c_s = -20$  and function  $f_{sph}$

$N$	$D$	$N_0$	$ D_{X,0} $	$\min_{X,0}$	$\max_{X,0}$	$\mu_{X,0}$	$\sigma_{X,0}^2$	$ D_{X,1} $
2	4	1	500	400	9 000	2 017.6	1 244 090	351
2	10	7	500	1 000	9 900	3 239.6	1 766 472	460
2	100	97	500	1 500	13 100	4 840.2	2 913 124	493
3	8	1	500	700	7 300	2 597.6	1 482 594	488
3	20	13	500	2 600	17 900	6 369.0	5 246 859	500
3	100	93	500	2 800	22 300	10 319.4	9 648 924	500
$N$	$D$	$N_0$	$ D_{X,1} $	$\min_{X,1}$	$\max_{X,1}$	$\mu_{X,1}$	$\sigma_{X,1}^2$	$ D_{X,2} $
2	4	1	351	0	3 600	862.7	518 636	225
2	10	7	460	0	10 500	1 054.3	1 030 785	359
2	100	97	493	0	6 900	1 080.5	773 576	457
3	8	1	488	0	6 200	748.0	715 611	430
3	20	13	500	0	9 700	1 554.8	2 139 917	500
3	100	93	500	0	12 800	2 087.6	4 274 366	500
$N$	$D$	$N_0$	$ D_{X,2} $	$\min_{X,2}$	$\max_{X,2}$	$\mu_{X,2}$	$\sigma_{X,2}^2$	$ D_{X,3} $
2	4	1	225	0	5 100	831.1	752 277	138
2	10	7	359	0	4 100	763.0	424 394	256
2	100	97	457	0	6 000	903.5	669 835	365
3	8	1	430	0	4 200	604.7	488 537	348
3	20	13	500	0	13 800	1 231.0	1 885 419	500
3	100	93	500	0	12 600	1 577.4	2 411 149	500



Table 6: Evaluation of data from Experiment 3 with  $c_0 = -40$ ,  $c_s = -20$  and function  $f_{hce}$

$N$	$D$	$N_0$	$ D_{X,0} $	$\min_{X,0}$	$\max_{X,0}$	$\mu_{X,0}$	$\sigma_{X,0}^2$	$ D_{X,1} $
2	4	1	500	400	6 400	1 948.8	1 016 179	368
2	10	7	500	1 000	7 800	3 191.4	1 570 546	458
2	100	97	500	1 700	9 500	4 481.4	2 077 034	487
3	8	1	500	400	7 700	2 520.6	1 554 436	483
3	20	13	500	1 700	15 100	6 523.2	5 161 462	500
3	100	93	500	3 400	20 800	9 861.0	8 493 979	500
$N$	$D$	$N_0$	$ D_{X,1} $	$\min_{X,1}$	$\max_{X,1}$	$\mu_{X,1}$	$\sigma_{X,1}^2$	$ D_{X,2} $
2	4	1	368	0	4 900	892.1	696 813	230
2	10	7	458	0	5 900	1 018.3	784 991	371
2	100	97	487	0	7 700	1 092.6	932 923	448
3	8	1	483	0	5 400	767.9	665 658	424
3	20	13	500	0	11 900	1 563.6	2 322 035	500
3	100	93	500	0	11 600	1 950.8	4 057 459	500
$N$	$D$	$N_0$	$ D_{X,2} $	$\min_{X,2}$	$\max_{X,2}$	$\mu_{X,2}$	$\sigma_{X,2}^2$	$ D_{X,3} $
2	4	1	230	0	5 800	764.8	612 455	132
2	10	7	371	0	5 700	827.2	585 970	272
2	100	97	448	0	5 900	989.3	720 108	360
3	8	1	424	0	5 300	612.5	509 301	368
3	20	13	500	0	10 200	1 159.6	1 652 248	500
3	100	93	500	0	13 600	1 478.8	2 400 471	500

Table 7: Evaluation of data from Experiment 3 with  $c_0 = -40$ ,  $c_s = -20$  and function  $f_{sch}$

$N$	$D$	$N_0$	$ D_{X,0} $	$\min_{X,0}$	$\max_{X,0}$	$\mu_{X,0}$	$\sigma_{X,0}^2$	$ D_{X,1} $
3	8	3	500	1 200	12 900	4 664.6	3 911 887	474
3	20	15	500	2 300	16 000	6 931.2	5 668 627	492
3	100	95	500	2 900	17 500	8 996.0	8 277 224	491
$N$	$D$	$N_0$	$ D_{X,1} $	$\min_{X,1}$	$\max_{X,1}$	$\mu_{X,1}$	$\sigma_{X,1}^2$	$ D_{X,2} $
3	8	3	474	0	12 600	1 171.7	2 251 437	371
3	20	15	492	0	15 000	1 270.7	2 516 623	452
3	100	95	491	0	14 100	1 365.4	2 824 870	446
$N$	$D$	$N_0$	$ D_{X,2} $	$\min_{X,2}$	$\max_{X,2}$	$\mu_{X,2}$	$\sigma_{X,2}^2$	$ D_{X,3} $
3	8	3	371	0	5 800	990.0	1 037 286	255
3	20	15	452	0	11 300	990.7	1 588 719	367
3	100	95	446	0	6 000	942.2	1 108 582	380

Table 8: Evaluation of data from Experiment 3 with  $c_0 = -40$ ,  $c_s = -20$  and function  $f_{dia}$

$N$	$D$	$N_0$	$ D_{X,0} $	$\min_{X,0}$	$\max_{X,0}$	$\mu_{X,0}$	$\sigma_{X,0}^2$	$ D_{X,1} $
3	3	1	500	700	13 300	2 557.0	1 845 011	1
3	10	8	500	2 300	26 200	5 722.8	9 167 760	1
4	3	1	500	1 500	28 500	6 533.4	13 461 824	46
4	10	8	500	6 200	36 900	14 990.2	26 613 084	75
$N$	$D$	$N_0$	$ D_{X,1} $	$\min_{X,1}$	$\max_{X,1}$	$\mu_{X,1}$	$\sigma_{X,1}^2$	$ D_{X,2} $
3	3	1	1	1 000	1 000	1 000.0	0	0
3	10	8	1	500	500	500.0	0	0
4	3	1	46	300	13 500	3 021.7	9 384 310	4
4	10	8	75	400	8 000	2 809.3	3 977 646	11
$N$	$D$	$N_0$	$ D_{X,2} $	$\min_{X,2}$	$\max_{X,2}$	$\mu_{X,2}$	$\sigma_{X,2}^2$	$ D_{X,3} $
3	3	1	0	$\infty$	$-\infty$	0.0	0	0
3	10	8	0	$\infty$	$-\infty$	0.0	0	0
4	3	1	4	1 400	16 600	6 825.0	33 761 875	0
4	10	8	11	900	13 100	3 800.0	11 474 545	0